

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Primož Kralj

Realizacija sistema poslovnega obveščanja v CPK d.d.

MAGISTRSKO DELO

Ljubljana, 2010

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Primož Kralj

Realizacija sistema poslovnega obveščanja v CPK d.d.

MAGISTRSKO DELO

Mentor: prof. dr. Viljan Mahnič

Ljubljana, 2010

Univerza
v Ljubljani

Fakulteta za računalništvo
in informatiko

Tržaška 25
1000 Ljubljana, Slovenija
telefon: 01 476 84 11
faks: 01 426 46 47
www.fri.uni-lj.si
e-mail: dekanat@fri.uni-lj.si

Št.: 100-MAG-ISO/2008

Datum: 15. 9. 2008



Primož KRALJ, univ. dipl. inž. rač. in inf.

Ljubljana

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

Naslov naloge: **Realizacija sistema poslovnega obveščanja v podjetju CPK, d.d.**

Implementation of a business intelligence system in CPK, d.d.

Tematika naloge:

Sistem poslovnega obveščanja predstavlja pomembno nadgradnjo celovitih programskih rešitev ERP, saj omogoča sprotno analizo doseženih rezultatov in spremljanje ključnih kazalnikov poslovanja. Zato je v mnogo podjetjih izgradnja tovrstnih sistemov ena najpomembnejših nalog na področju informatike.

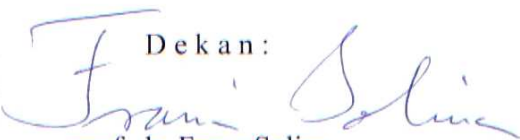
V magistrski nalogi predstavite vpeljavo sistema poslovnega obveščanja v podjetju CPK Koper. V ta namen definirajte strukturo podatkovnega skladišča, opredelite potrebne podatkovne vire in realizirajte postopke za pripravo in polnjenje podatkov. Podatke iz podatkovnega skladišča nato uporabite za izdelavo ustreznih poročil in nadzornih plošč. Za realizacijo celotnega sistema definirajte ustrezno arhitekturo in utemeljite izbor najprimernejših orodij. V nalogi podrobno opišite realizacijo celotnega sistema – od predstavitve podatkovnih virov do analiz in poročil, ki so na voljo končnim uporabnikom.

Mentor:


prof. dr. Viljan Mahnič



Dekan:


prof. dr. Franc Solina

IZJAVA O AVTORSTVU

magistrskega dela

Spodaj podpisani/-a Primož Kralj,

z vpisno številko 63970080,

sem avtor/-ica magistrskega dela z naslovom

Realizacija sistema poslovnega obveščanja v CPK d.d.

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelal/-a samostojno pod vodstvom mentorja (naziv, ime in priimek)

prof. dr. Viljan Mahnič

in somentorstvom (naziv, ime in priimek)

- so elektronska oblika magistrskega dela, naslova (slov., angl.), povzetka (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela
- in soglašam z javno objavo elektronske oblike magistrskega dela v zbirki »Dela FRI«.

V Ljubljani, dne 30.7.2010

Podpis avtorja/-ice:



Zahvala

V prvi vrsti se zahvaljujem prof. dr. Viljan Mahničju za številne strokovne nasvete in priporočila pri izdelavi magistrskega dela. Zahvaljujem se vsem sodelavcem na CPK, ki so bili v izdelavo magistrskega dela posredno ali neposredno vpleteni ter vodstvu združbe za nudeno podporo. Zahvala gre tudi prijateljem, ki so me s pridobivanjem lastnih akademskih nazivov, spodbujali pri izdelavi magistrskega dela.

Posebno zahvalo namenjam mami Mojci za spodbujanje na moji učni poti ter očetu Dimitriju, ki me je že v otroštvu naučil logičnega razmišljanja in mi tako omogočil uspehe, med katerimi je tudi dokončanje magistrskega dela.

Kazalo

Povzetek	1
Abstract	2
1. Uvod	3
2. Zgodovina sistemov poslovnega obveščanja	4
3. Osnove podatkovnih skladišč	7
4. Področna podatkovna skladišča	9
4.1. Dimenzijski podatkovni model	9
4.1.1. Tabele dejstev	9
4.1.2. Dimenzijske tabele	10
4.1.3. Različice dimenzijskega podatkovnega modela	11
4.2. Tehnološki vidik področnih podatkovnih skladišč	13
4.2.1. Tehnologija shranjevanja podatkov	13
4.2.2. Agregati	14
4.2.3. Paralelizem in aktivni predpomnilnik	15
5. Proces ETL	16
5.1. Shramba vmesnih podatkov	16
5.2. Aktivnosti procesa ETL	17
5.2.1. Izbor podatkov	17
5.2.2. Čiščenje podatkov	20
5.2.3. Poenotenje podatkov	23
5.2.4. Dostava podatkov	24
6. Orodja poslovnega obveščanja	27
6.1. Uporabniki orodij poslovnega obveščanja	29
7. Pristop življenjskega cikla	30
7.1. Načrtovanje in vodenje projekta	31
7.2. Določitev potreb združbe	32
7.3. Tehnološka sled	33
7.4. Podatkovna sled	34
7.4.1. Priprava dimenzijskega podatkovnega modela	35
7.4.2. Določitev fizične zasnove	37
7.4.3. Načrtovanje in realizacija procesa ETL	38
7.5. Aplikativna sled	39
7.6. Namestitev sistema	39
7.7. Vzdrževanje in nadgradnje sistema	40

8. Podatkovno skladišče CPK d.d.	42
8.1. <i>Načrtovanje in vodenje projekta</i>	42
8.2. <i>Določitev potreb združbe</i>	43
8.3. <i>Določitev tehnološke zasnove in platforme</i>	44
8.4. <i>Načrtovanje in realizacija uporabniškega okolja</i>	46
8.5. <i>Priprava matrike podatkovnega skladišča</i>	46
9. Področno podatkovno skladišče Glavna knjiga	47
9.1. <i>Izvorni sistem</i>	47
9.2. <i>Priprava dimenzijskega podatkovnega modela</i>	49
9.2.1. Pregled dimenzijskih tabel	50
9.3. <i>Načrtovanje in realizacija procesa ETL</i>	54
9.3.1. Globalni paket	54
9.3.2. Gradniki opravil	55
9.3.3. Opravilo Polnjenje dimenzije Konto GK	56
9.3.4. Opravilo Polnjenje dimenzije SM	58
9.3.5. Opravilo Polnjenje dimenzije SN	59
9.3.6. Opravilo Polnjenje dimenzij Stik	60
9.3.7. Opravilo Polnjenje tabele dejstev Postavka GK	61
9.4. <i>Določitev fizične zasnove</i>	62
9.4.1. Določitev standardov za imenovanje objektov	62
9.4.2. Načrtovanje fizičnega modela	63
9.4.3. Načrtovanje agregatov in indeksov	63
9.4.4. Določitev tehničnih lastnosti podatkovne baze	64
9.5. <i>Snovanje kocke OLAP</i>	66
9.5.1. Merljiva dejstva	66
9.5.2. Uporaba dimenzij	66
9.5.3. Izračunana dejstva	68
9.5.4. Ključni kazalniki poslovanja	69
9.6. <i>Primeri uporabe</i>	72
10. Preostala področna podatkovna skladišča	74
10.1. <i>Področno podatkovno skladišče Saldakonti</i>	74
10.1.1. Dimenzijski podatkovni model	74
10.1.2. Predstavitev specifičnih rešitev	75
10.1.3. Primeri uporabe	77
10.2. <i>Področno podatkovno skladišče Materialno poslovanje</i>	78
10.2.1. Dimenzijski podatkovni model	78
10.2.2. Predstavitev specifičnih rešitev	78

10.2.3. Primeri uporabe	79
10.3. Področno podatkovno skladišče Evidenca dela	80
10.3.1. Dimenzijski podatkovni model	80
10.3.2. Predstavitev specifičnih rešitev	81
10.3.3. Primeri uporabe	81
11. Namestitev sistema	82
11.1. Namestitev zalednega sistema.....	82
11.2. Namestitev čelnega sistema	83
12. Vzdrževanje in nadgradnje.....	84
13. Sklep.....	85
Viri	87

Slovar izrazov

slovensko	kratica	angleško
celo število		integer
celovita programska rešitev	ERP	enterprise resource planning
čelni sistem		front room
degenerirana dimenzija		degenerate dimension
dimenzija revizije		audit dimension
dimenzijski podatkovni model		dimensional data model
ekstrapolacija		extrapolation / forecasting
izbor, čiščenje, poenotenje, dostava	ECCD	extract, clean, conform, deliver
izbor, preoblikovanje, nalaganje	ETL	extract, transform, load
izračunana dejstva		calculated members
izraz		expression
ključni kazalnik poslovanja	KPI	key performance indicator
kocka OLAP		OLAP cube
kumulativni posnetek		accumulating snapshot
matrika podatkovnega skladišča		data warehouse bus matrix
nadomestni ključ		surrogate key
nadzorna plošča		performance dashboard
napovedovalno modeliranje		predictive modeling
ničelna vrednost		null value
opozorila		alerts, alarms
obstojna shramba vmesnih podatkov		persistent staging area
odjemalec		client
optimizacija poslovanja		business optimization
orodja poslovnega obveščanja		business intelligence tools
particija		partition
periodični posnetek		periodic snapshot
podatkovno rudarjenje		data mining
podatkovno skladišče		data warehouse
podatkovno skladišče združbe		enterprise data warehouse
področje preoblikovanja podatkov		staging area
področno podatkovno skladišče		data mart
poizvedba		query
poslovno analiziranje		business analytics
poslovno obveščanje	BI	business intelligence
posnetek zaslona		screenshot
potrošniki podatkov		information consumers
pregled izvornih sistemov		source system tracking report
pregled toka podatkov		logical data map
prehodna shramba vmesnih podatkov		transient staging area
prenova poslovnih procesov	BPR	business process reengineering
primarni ključ		primary key
razporejevalnik		scheduler
shramba vmesnih podatkov		staging area storage
sistem za upravljanje podatkovnih baz	SUPB	database management system
skladne dimenzije		conformed dimensions
skupina merljivih dejstev		measure group

slovensko	kratica	angleško
snežinkasta shema	SQL	snowflake schema
statistična analiza		statistical analysis
strežnik		server
strukturiran poizvedovalni jezik		structured query language
tabela napak		error event table
tuji ključ		foreign key
ustvarjalci podatkov		information producers
vrtanje v globino		drill down
vrtanje v širino		drill across
vsebnik		container
zaledni sistem		back room
zrno		grain
zvezdna shema		star join schema

Povzetek

Magistrsko delo z naslovom Realizacija sistema poslovnega obveščanja v CPK d.d. vsebuje predstavitev teoretičnega ozadja sistemov poslovnega obveščanja in praktični prikaz realizacije sistema.

Uvod v teoretični del naloge predstavlja zgodovinski pregled sistemov poslovnega obveščanja, kateremu sledi podroben opis posameznih elementov podatkovnega skladišča. Predstavljene so najpogostejše uporabljene tehnologije, različni pristopi k realizaciji ter najnovejši trendi na področju sistemov poslovnega obveščanja. Poseben poudarek je namenjen problematiki načina predstavitve informacij poslovnim uporabnikom, ki jih je zaradi različnih znanj in zahtev potrebno obravnavati individualno. Individualni pristop je mogoče zagotoviti z izbiro primernih orodij poslovnega obveščanja, med katerimi velja izpostaviti v zadnjem času še posebej popularne nadzorne plošče.

V nadaljevanju teoretičnega dela sledi podrobna predstavitev metodologije Življenjskega cikla, ki zagotavlja visoko stopnjo uspešnosti realizacije projekta. Metodologija sestoji iz zaporedja aktivnosti, ki zajemajo načrtovanje, izvedbo in namestitev sistema poslovnega obveščanja.

V praktičnem delu je predstavljen potek realizacije sistema poslovnega obveščanja CPK d.d. Realizacija v veliki meri sledi smernicam metodologije Življenjskega cikla, vendar je bila zaradi manjšega obsega projekta uporabljena metodologija poenostavljena. Kot posledica uporabe moderne platforme, pa je bilo potrebno metodologijo z vključitvijo dodatne aktivnosti celo nekoliko razširiti. Metodologija namreč ne predvideva določenih opravil, ki jih je pri realizaciji modernih sistemov poslovnega obveščanja potrebno izvesti.

Rezultat magistrskega dela predstavlja realiziran sistem poslovnega obveščanja. Sistem je predstavljen z opisom izbrane platforme, uporabljenih tehnologij, procesa ETL ter področnih podatkovnih skladišč. Pri opisu slednjih je poseben poudarek namenjen obrazložitvi posebnosti pri realizaciji podatkovnih transformacij in snovanju dimenzijskega podatkovnega modela. Opisani so načini realizacije izračunanih vrednosti in ključnih kazalnikov poslovanja. Za boljše razumevanje vsebine posameznih področnih podatkovnih skladišč so v zaključku prikazani primeri konkretnih poročil, katerih končni namen je zagotovitev kakovostnih informacij za učinkovito poslovno odločanje.

Ključne besede:

Poslovno obveščanje, podatkovno skladišče, proces ETL, ključni kazalnik poslovanja, dimenzijski podatkovni model.

Abstract

The thesis Implementation of a business intelligence system in CPK d.d. contains theoretical background regarding business intelligence systems and practical presentation of business intelligence system development in CPK d.d.

The introduction to the theoretical part of the thesis contains a historical review of the business intelligence systems followed by a detailed presentation of individual data warehouse elements. Common technologies, different approaches to realization and modern trends in business intelligence systems are presented. Special attention is given to the methods of presenting results to the business users, which have to be treated individually. Individual approach is possible due to variety of business intelligence tools, especially nowadays popular performance dashboards.

The theoretical part of the thesis continues with a detailed presentation of the Data Warehouse Lifecycle methodology, which ensures high project success rate. Methodology consists of task sequences, required for effective data warehouse design, development and deployment.

The realization of business intelligence system in CPK d.d. follows the guidelines provided by Data Warehouse Lifecycle methodology. However, due to limited project extent the applied methodology has been simplified. As a consequence of modern platform usage, which required additional system development steps, the methodology had to be extended by introduction of an additional task.

The result of the thesis is reflected by successful deployment of CPK's business intelligence system. The system is presented by description of the selected platform, ETL process and data marts. The developed data marts are covered individually, focusing on data transformation and dimensional data models specificities. Furthermore, calculated members and key performance indicators are covered in detail. In pursuit of providing better understanding of individual data marts content, examples of reports are depicted. Their final goal is delivery of quality information for effective business decision making.

Key words:

Business intelligence, data warehouse, ETL process, key performance indicator, dimensional data model.

1. Uvod

V vsaki združbi se pri izvajanju poslovnih procesov zbirajo, obdelujejo in hranijo podatki. Za ta namen so v uporabi t.i. transakcijski sistemi, katerih naprednejše različice, kot so na primer celovite programske rešitve ERP, uporabnikom omogočajo tudi določen nivo analiziranja podatkov. Z namenom dviga omenjenega nivoja na višjo raven, so se kot nadgradnja ERP sistemov za področje analiziranja podatkov, v drugi polovici 90. let pojavili prvi sistemi poslovnega obveščanja.

Realizacija sistema poslovnega obveščanja je v današnjih časih v številnih združbah ena izmed najpomembnejših nalog na področju informatike. Mnoge združbe si skušajo zagotoviti konkurenčno prednost z učinkovitim zbiranjem in analiziranjem podatkov ter odločanjem na podlagi kvalitetnih informacij. Po mednarodni raziskavi skupine Gartner iz leta 2008 [18], v kateri je sodelovalo 1500 direktorjev informatike, so po prioriteti sistemi poslovnega obveščanja uvrščeni na prvo mesto.

Osnovni cilj magistrskega dela je realizacija sistema poslovnega obveščanja, ki bo nosilec odločanja zagotavljal kakovostnejše informacije, kot jih trenutno nudi celovita programska rešitev ERP združbe. Ključni dejavniki uspeha, ki smo jih v CPK d.d. zagotovili še pred začetkom projekta, so:

- podpora vodstva
- zagotovljeno financiranje
- pripravljenost in sposobnost uporabe sistema s strani poslovnih uporabnikov
- skladnost ciljev sistema poslovnega obveščanja s strategijo združbe
- obstoj kakovostnih izvornih podatkov

Pri realizaciji sistemov poslovnega obveščanja se praviloma največ težav pojavi v procesu preoblikovanja podatkov iz izvornih sistemov v obliko, ki je primerna za sisteme poslovnega obveščanja. Za omenjeni proces, ki ga imenujemo proces ETL, se v povprečju potroši 70% časa celotne realizacije sistema poslovnega obveščanja [12], pri čemer vzrok težav najpogosteje predstavljajo neakovostni izvorni podatki. V primeru, da se pred začetkom realizacije ne opravi temeljite analize kakovosti izvornih podatkov, lahko število in kritičnost težav krepko preseže pričakovanja, kar podaljša trajanje celotnega projekta in posledično poveča stroške. Ker bo v našem primeru izvorni sistem skoraj v celoti predstavljal ERP sistem združbe, ki že v osnovi zagotavlja določeno stopnjo kakovosti podatkov, predvidevamo, da se bomo z omenjenimi težavami lahko uspešno soočili.

Med celotnim potekom projekta je ključnega pomena tesno sodelovanje med informatiko in poslovnimi uporabniki [22]. Sodelovanje zagotavlja, da bo sistem poslovnega obveščanja v zadovoljivi meri zadostil dejanskim potrebam nosilcev odločanja, kar posledično poveča verjetnost uspešnosti celotnega projekta.

Uspešna realizacija sistema poslovnega obveščanja združbi prinaša številne koristi. Te segajo od lažje merljivih z lokalnim učinkom, do težje merljivih, ki učinkujejo na združbo v celoti. V prvo skupino uvrščamo koristi, kot sta na primer skrajšanje potrebnega časa pri pripravi novih poročil v oddelku za informatiko in prihranek časa pridobivanju informacij s strani končnih uporabnikov v posameznih oddelkih združbe. V skupini težje merljivih koristi pa naštejmo pridobitve kot so, povečanje količine in kakovosti informacij, učinkovitejše poslovne odločitve in izboljšanje poslovnih procesov [24].

2. Zgodovina sistemov poslovnega obveščanja

Na svetovnem spletu je moč najti vrsto definicij za sisteme poslovnega obveščanja. Pri iskanju naše definicije želimo odgovoriti na dve temeljni vprašanji. Kaj so sistemi poslovnega obveščanja in čemu so namenjeni?

Sistemi za poslovno obveščanje so se z razvojem novih tehnologij skozi daljša časovna obdobja močno spreminjali, njihov namen pa ves čas ostaja enak, kar poenostavi odgovor na drugi del našega vprašanja. Namen sistemov za poslovno obveščanje je zagotavljanje informacij za poslovne odločitve. Za odgovor na prvi del vprašanja, pa si bomo v nadaljevanju pogledali razvoj sistemov poslovnega obveščanja skozi čas.

Ker dvomimo, da bi se nosilci poslovnih odločitev odločali le na podlagi intuicije, lahko trdimo, da sistemi za poslovno obveščanje v svoji primitivni obliki obstajajo od nastanka prvih združb in posledično prvih poslovnih odločitev.

Pred obdobjem računalništva so sisteme za poslovno obveščanje predstavljali ljudje - analitiki, ki so imeli znanja, da so iz množice zapisov o poslovnih dogodkih z matematičnimi in statističnimi metodami skušali odgovoriti na vprašanja, ki so jim jih zastavili nosilci poslovnih odločitev. Ta po vsebini najverjetneje niso bila zelo različna od vprašanj, ki si jih zastavljajo današnji ravnatelji združb, saj jih je ravno tako zanimalo »kateri izdelek se najboljšje prodaja?« ali »kateri kupci najmanj zamujajo s plačili?«. Le čas potreben za odgovor je bil tedaj neprimerljivo daljši.

Masovno zbiranje podatkov v digitalni obliki se je začelo v 60. letih s prihodom mrežnega in hierarhičnega podatkovnega modela. Leta 1970 je Edgar F. Codd predstavil delo z naslovom »Relacijski podatkovni model za upravljanje velikih baz podatkov« [2], ki je postavil temelje za podatkovne baze, kot jih poznamo še danes. Obstoje teorije je v samo nekaj letih sprožil množičen prihod relacijskih sistemov za upravljanje s podatkovnimi bazami (SUPB) kot so *Ingres*, *SyBase*, *System R* in kasneje *DB2*. Posledično so se na tržišču pojavila tudi prva orodja, namenjena analiziranju podatkov in izdelavi poročil, ki pa jih pestijo mnoge težave. Prva težava je bila v tem, da analitik pri svojem delu ni bil samostojen, saj je za prenos podatkov iz podatkovne baze potreboval pomoč programerjev. Po prenosu podatkov, ki so bili od tega trenutka dalje neusklajeni z osnovnim virom, pa je veliko težavo predstavljala omejenost količine podatkov, nad katero so ta orodja lahko operirala. Ta je namreč le redko zadostovala dejanskim potrebam. Uporabniški vmesnik na nivoju ukazne vrstice in velika neodpornost na vsebinske anomalije v podatkih, kot so manjkajoče vrednosti ali delni zapisi, sta poleg že omenjene pomnilniške omejenosti pripomogla k temu, da se orodja za analizo v tistih časih še niso množično razširila [1].

S prihodom prvih osebnih računalnikov v začetku 80. let se je svet analiziranja podatkov hitro začel spreminjati. Za njihov bliskovito širitev v poslovne vode je najzaslužnejši Daniel Bricklin, ki je leta 1978 na univerzi Harvard začel s projektom VisiCalc. Ideja njegovega programa je bila v tem, da naj bi upravljali s številkami na tako enostaven način, kot so to obstoječi programi počeli z besedili. Leta 1979 je pod okriljem združbe *Software Arts* tako nastal prvi program za računske preglednice in učinek na poslovni svet je bil enormen. Opravila, za katera so v finančnih službah porabili ure dela, so bila z *VisiCalcom* opravljena v minutah [3]. *Apple II* je z *VisiCalcom* postal sinonim za poslovni računalnik in njegova prodaja je bliskovito naraščala. Kot zanimivost omenimo še, da Bricklin programa *VisiCalc* ni

patentiral, kar je združbi *Lotus* omogočilo, da je leta 1983 brez težav razvila svetovno znano aplikacijo *Lotus 1-2-3*.

Od tedaj dalje analitikom ni bilo več potrebno opravljati svojega dela na velikih računalnikih v informacijskih centrih združbe, kar je od njih zahtevalo tudi določena tehnična znanja, ampak so to lahko počeli na uporabniku prijaznejših osebnih računalnikih v zasebnosti lastnih pisarn. Ostala je le še ena večja težava. Za pridobitev novih podatkov so se še vedno morali obrniti na programerje v informacijskih centrih, kar je bilo nepraktično in počasno.

Rešitev težave se je pojavila v poznih 80. letih s prvimi relacijskimi sistemi za upravljanje s podatkovnimi bazami na osnovi dvonivojske arhitekture odjemalec – strežnik, ob boku katerih so se pojavila mnoga orodja za analiziranje podatkov. Čeprav je večina izmed njih za poizvedovalni jezik uporabljala SQL, je šele standardizacija jezika v letu 1989 omogočila enostavne prehode med različnimi ponudniki SUPB-jev. Uvedbo novih sistemov so najbolj zavirali visoki stroški. Poleg stroškov strojne, programske in omrežne opreme je bil prehod priložnost za prenovo poslovnih procesov, kar je povzročilo še dodatne stroške za svetovalne storitve združb s tovrstnimi izkušnjami. Kljub temu pa so množični prehodi na nove sisteme in prenove poslovnih procesov poganjali informacijsko industrijo. V začetku 90. let se pojavijo prve celovite programske rešitve (ERP). Tako poimenovanje je začel prvi uporabljati *Gartner Group* za sisteme združb kot so npr. *SAP*, *Baan*, *Peoplesoft*, ki so poleg proizvodnje pokrivali tudi ostale poslovne funkcije [9]. Ti sistemi so analitikom bistveno olajšali delo, saj so od tedaj lahko večino podatkov o poslovanju črpali iz enega vira.

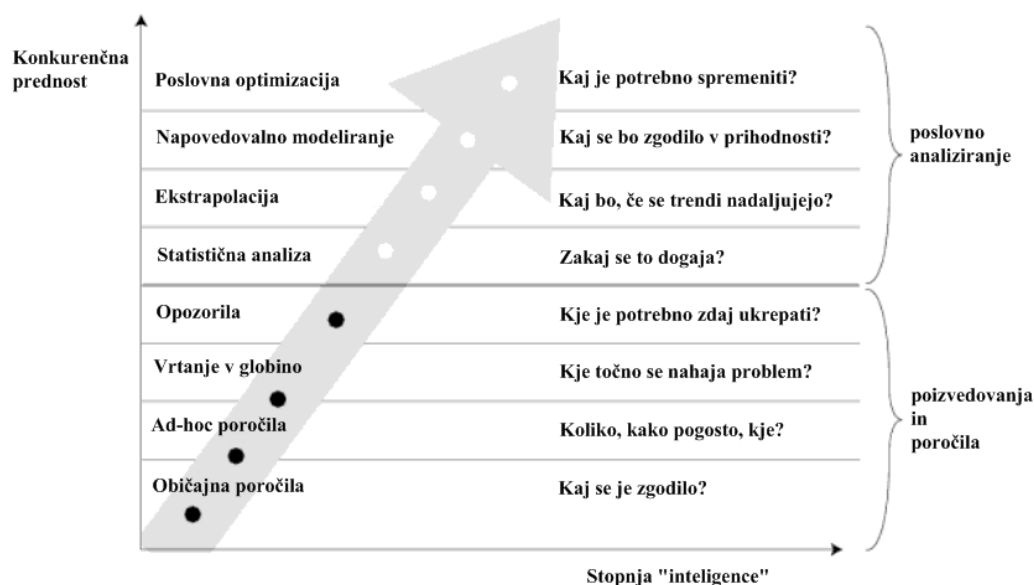
Skozi zgodovinski pregled smo prišli v čas druge polovice 90. let, ko se je začel množično uporabljati naš iskani pojem poslovne inteligence oz. poslovnega obveščanja. Definirali ga bomo kot široko množico aplikacij in tehnologij za zbiranje, shranjevanje in analiziranje podatkov s ciljem zagotoviti informacije nosilcem odločanja [20]. Kot je razvidno iz definicije, so že prvi ERP sistemi vsebovali določene prvine sistemov poslovnega obveščanja. Preko vnaprej pripravljenih poročil so bili namreč sposobni odgovoriti na mnoga specifična vprašanja. Izkazalo se je, da je potrebno za nadaljnji korak, ki bi omogočal odgovore na širšo množico t.i. ad-hoc vprašanj in uporabo ostalih tehnologij poslovnega obveščanja, podatke iz relacijske podatkovne baze združbe preurediti.

Čeprav se pojem podatkovnega skladišča, ki je osnova za moderne sisteme poslovnega obveščanja in z realizacijo, katerega se bomo ukvarjali večjem delu te naloge, prvič omenja že leta 1988, pa se taki sistemi množično razširijo šele v drugi polovici 90 let. Najpomembnejša guruja na tem področju sta Bill Inmon, ki leta 1991 izda knjigo *Building the Data Warehouse* [11] in Ralph Kimball s knjigo *The Data Warehouse ToolKit*, katere prva izdaja izide leta 1996, druga pa leta 2002 [14]. O podatkovnih skladiščih, ki ne predstavljajo zgolj novega načina shranjevanja podatkov, bomo še veliko povedali v nadaljevanju, tako da se sedaj lahko osredotočimo na tehnologije, ki so se na osnovi podatkovnih skladišč razvile.

Tehnologije poslovnega obveščanja delimo na običajna poročila, ad-hoc poročila, vrtanje v globino, opozorila ter orodja poslovnega analiziranja [5]. S stališča uporabniške izkušnje prva tehnologija ni nova, saj so sorodno funkcionalnost omogočale že aplikacije, ki so delovale nad relacijskimi podatkovnimi bazami in kasneje ERP sistemi. Nasprotno pa nam tehnologije poizvedovanj in poročil nad podatkovnim skladiščem omogočajo večjo fleksibilnost ter hitre odgovore na poljubna ad-hoc vprašanja o poslovnih dogodkih iz preteklosti.

Poslovno analiziranje, pod katerim razumemo orodja za kompleksnejše statistično in napovedovalno modeliranje, kot je na primer podatkovno rudarjenje, pa za razliko od poizvedovanj in poročil ne išče odgovora na to, kaj se zgodilo v preteklosti, ampak skuša dogajanje obrazložiti in napovedati prihodnost. Z omenjeno tehnologijo je moč v podatkih poiskati vzorce, trende in korelacije in tako nosilcu odločanja podati nova znanja ter olajšati poslovne odločitve [5, 19].

Na sliki 1 so navedene tehnologije poslovnega obveščanja prikazane glede na stopnjo »intelligence« in konkurenčne prednosti, ki jo zagotavljajo.



Slika 1. Tehnologije poslovnega obveščanja

Z razširitvijo kroga uporabnikov sistemov poslovnega obveščanja tudi na vrhnje ravnateljstvo se je v modernejših sistemih uveljavil koncept ključnih kazalnikov poslovanja, preko katerih ugotavljamo uspešnost združbe ali posameznika na različnih nivojih. Te pogosto prikazujemo na nadzornih ploščah, ki nosilcu odločanja predstavljajo vir, ki je s stališča uporabe izredno enostaven, a ima lahko visoko informacijsko vrednost.

Vedno zmogljivejša strojna oprema je v zadnjih letih omogočila nov pristop realizacije podatkovnih skladišč, ki niso polnjena periodično, ampak vsebujejo »žive« podatke. Take potrebe imajo predvsem združbe, ki sisteme poslovnega obveščanja uporabljajo za odločanje na nivojih, kjer npr. dnevni zamik podatkov ne zadostuje. Kot primer navedimo področje prodaje velike večnacionalne združbe, kjer se podatkovno skladišče uporablja za pripravo individualnih promocij [12].

Čeprav se v nalogi z njimi ne bomo srečali, pa za konec omenimo še tehnologije, ki jih razumemo pod terminom poslovno obveščanje 2.0 in predstavljajo odmik od klasičnega pristopa, kjer vir podatkov predstavlja zgolj podatkovno skladišče. Taki sistemi naj bi vključevali mnoge heterogene vire, kot so blogi, elektronska pošta in spletne strani ter bi preko semantičnih oznak omogočali njihovo avtomatsko integracijo. Preko spletnega uporabniškega vmesnika bi bilo uporabniku enostavno poiskati odgovor na želeno vprašanje brez poznavanja strukture podatkov ali pomoči informatikov. Velja poudariti, da te tehnologije ne nadomeščajo »klasičnega« pristopa, ampak ga le dopolnjujejo [16].

3. Osnove podatkovnih skladišč

Na svetovnem spletu je moč najti mnogo definicij za pojem podatkovnega skladišča. Za potrebe te naloge se bomo oprli na definiciji dveh, v prejšnjem poglavju omenjenih gurujev tega področja. Bill Inmon podatkovno skladišče definira kot pojmovno in časovno naravnano, poenoteno in prirastkovno zbirko podatkov, ki podpira poslovno odločanje [11].

Pojmovna naravnanoost se nanaša na organiziranost podatkovnega skladišča in je nasprotje t.i. procesni naravnanoosti, ki je značilna za ERP in druge transakcijske informacijske sisteme. Področja ERP sistema predstavljajo poslovni procesi, kot so npr. nabava in prodaja, podatkovno skladišče pa je po Inmonu urejeno glede na poslovne subjekte, kot so delavec, artikel, dobavitelj ipd.

Časovna naravnanoost podatkovnega skladišča je načelo, ki veli, da naj podatkovno skladišče vsebuje podatke o daljšem časovnem obdobju, ki so praviloma periodični posnetki stanja osnovnega podatkovnega vira.

Poenotenje podatkovnega skladišča pomeni, da so podatki v podatkovnem skladišču vsebinsko usklajeni ne glede na to, da so črpani iz heterogenih virov, v katerih je isti vsebinski koncept lahko različno poimenovan.

Zadnja lastnost podatkovnega skladišča po Inmonu je prirastkovnost, ki pomeni, da se podatki v podatkovnem skladišču ne spreminjajo, ampak le periodično dodajajo ter seveda berejo s strani uporabnikov.

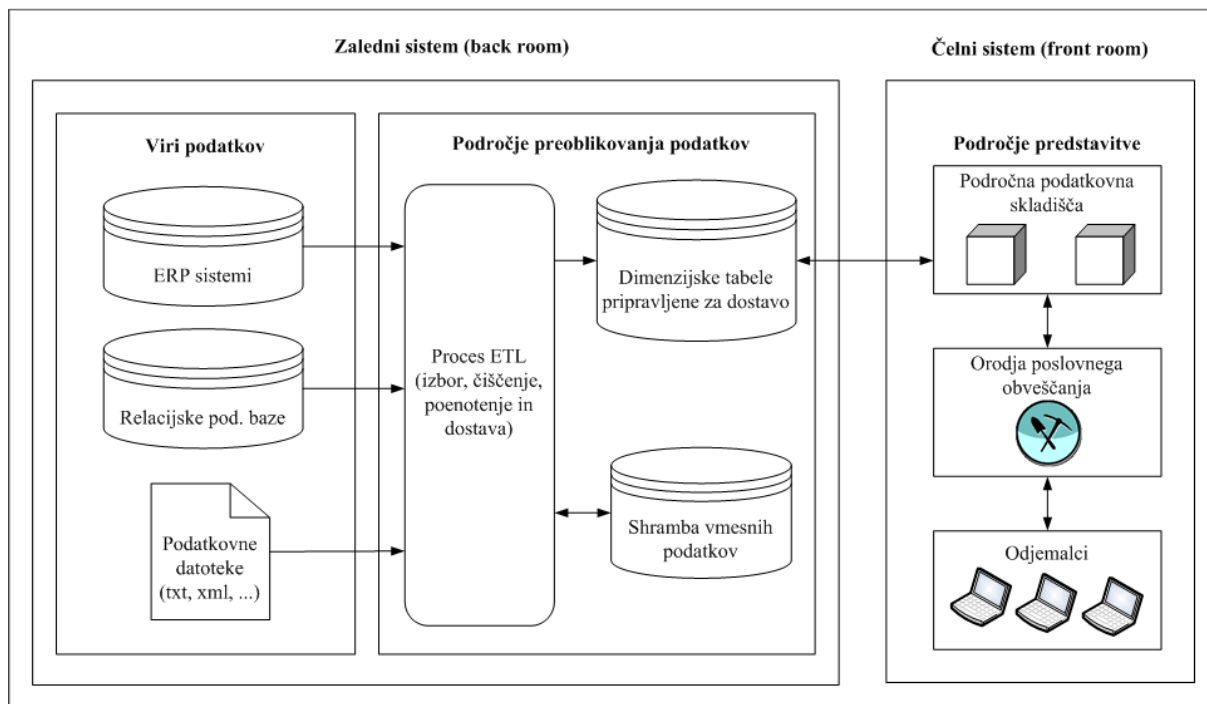
Bistveno drugače si je definicijo podatkovnega skladišča zamislil Ralph Kimball [12]. Ta se glasi: »Podatkovno skladišče je sistem, ki izbere, prečisti, poenoti in dostavi podatke iz podatkovnih virov v dimenzijski podatkovni model in nato omogoča poizvedovanja in analizo z namenom podpore poslovnemu odločanju.«

Pri primerjavi definicij lahko hitro opazimo, da obe kot smoter podatkovnega skladišča navajata podporo odločanju. Tu pa se podobnost konča, saj Kimball v ospredje definicije postavlja proces polnjenja podatkovnega skladišča (ETL), medtem ko Inmon poudarja lastnosti samega skladišča. Po Kimbalovi definiciji je v sistem podatkovnega skladišča nedvomno vključen tudi proces ETL, kar iz Inmonove definicije ni razvidno.

Poudarjanje procesa polnjenja, ki sestoji iz podprocesov izbora, čiščenja, poenotenja in dostave podatkov v dimenzijski podatkovni model, lahko obrazložimo s Kimballovo oceno, da se pri projektu realizacije podatkovnega skladišča zanj potroši v povprečju 70% časa. Pri tem je potrebno opozoriti, da gre pri realizaciji podatkovnega skladišča tipično za več ločenih projektov realizacije področnih podatkovnih skladišč z lastnimi resursi in časovnimi okviri. Na ta način se podatkovno skladišče s časom dopolnjuje, njegova funkcionalnost pa širi. Pravilneje je torej, da realizacijo podatkovnega skladišča v celoti razumemo kot proces in ne projekt.

Pojasnimo, da smo se pri ugotavljanju razlik do sedaj osredotočili izključno na razlike v definicijah podatkovnega skladišča. Najpomembnejša vsebinska razlika med gurujema je namreč v pristopu k realizaciji podatkovnega skladišča, ki pa jo bomo podrobneje obrazložili

v pod poglavju 7.4. Naj pa ne bo odveč, da že na tem mestu razjasnimo, da bomo pri realizaciji sistema za poslovno obveščanje CPK d.d. v veliki meri sledili Kimballovemu pristopu.



Slika 2. Struktura podatkovnega skladišča

Kot je razvidno iz slike 2, je podatkovno skladišče po Kimballu tako široko zastavljen pojem, da ga lahko enačimo s sistemom poslovnega obveščanja. Na najvišjem nivoju je razdeljen na čelni sistem (front room) in zaledni sistem (back room). Slednji poleg procesa polnjenja zajema še vire podatkov in shrambo vmesnih podatkov. Pomembno je, da končnim uporabnikom vanjo neposredno ni dovoljeno dostopati, saj se v njej izvajajo procesi podatkovnih transformacij.

Končni produkt teh transformacij, oziroma procesa ETL so podatki, ki so urejeni na način, da ustrezajo dimenzijskemu podatkovnemu modelu ter kot taki predstavljajo vir področnim podatkovnim skladiščem. V primeru sistema s tronivojsko arhitekturo uporabniki do njih dostopajo preko aplikacijskega strežnika, na katerem so nameščena orodja poslovnega obveščanja, kot so poizvedovanja, poročila in poslovno analiziranje ter si tako zagotovijo informacije za odločanje.

S tako strukturo podatkovnega skladišča je uporabniku zagotovljeno, da dostopa izključno do podatkov, ki kot končni produkt procesa ETL predstavljajo eno in edino inačico resnice. Na ta način je tudi proces odločanja lahko kakovostnejši, ta pa kot rečeno predstavlja končni namen podatkovnega skladišča oziroma sistema podatkovnega obveščanja.

V naslednjih poglavjih bomo podrobneje predstavili posamezne elemente podatkovnega skladišča.

4. Področna podatkovna skladišča

Po Kimballu je področno podatkovno skladišče množica dimenzijskih tabel in tabel dejstev, ki zajemajo določeno poslovno področje [12]. Govora je torej o množici podatkov, urejenih v obliki, ki ustrezajo dimenzijskemu podatkovnemu modelu ter vsebinsko pokrivajo določeno poslovno področje, kot je na primer prodaja, materialno poslovanje, glavna knjiga ipd.

Področna podatkovna skladišča imajo naslednje lastnosti:

1. Podatke črpajo neposredno iz tabel podatkovnega vira in ne iz oddelčnih pogledov nad temi podatki.
2. Zajemajo dovolj podrobne podatke, da je omogočeno vrtanje v globino do najnižjega nivoja.
3. Realizacija posameznega izmed njih je neodvisna od realizacije ostalih.

Zadnja točka zahteva dodatno razlago. Področna podatkovna skladišča so osredotočena na oddelke ali skupine uporabnikov z lastnimi potrebami poslovnega obveščanja. Ta osredotočenost se kaže v optimizaciji posameznega področnega podatkovnega skladišča za potrebe vsake izmed skupin uporabnikov. Taka zasnova pa ima pomanjkljivost, saj nimamo celovitega pogleda na podatke vseh področnih podatkovnih skladišč skupaj. S tem namenom se je skozi leta razvil koncept t.i. skladnih dimenzij, ki povezujejo področna podatkovna skladišča v celoto [10, 14]. Realizacija posameznih področnih podatkovnih skladišč je torej neodvisna le do določene mere, saj je za moderen sistem poslovnega obveščanja s skupnimi dimenzijami potrebno upoštevati, da se področna podatkovna skladišča med seboj prekrivajo. Koncept skladnih dimenzij bomo podrobneje predstavili v nadaljevanju.

4.1. Dimenzijski podatkovni model

Dimenzijski podatkovni model nima enega avtorja, ampak je posledica strmenja mnogih oseb k podatkovni strukturi, ki bi bila visoko zmogljiva za potrebe podatkovnega skladišča in bila hkrati enostavna za modeliranje in razumevanje [14].

V dimenzijskem podatkovnem modelu se srečamo z dvema vrstama tabel. Tabele dejstev vsebujejo numerične vrednosti oz. meritve, dimenzijske tabele pa te meritve dodatno opisujejo oz. jih umeščajo v kontekst.

4.1.1. Tabele dejstev

Tabele dejstev predstavljajo središče dimenzijskega podatkovnega modela in so praviloma največje tabele izmed vseh. Njihova velikost je odvisna predvsem od izbora t.i. zrna, ki določa, katera meritev oz. transakcija predstavlja posamezen zapis v tabeli dejstev. V modernih sistemih, podprtih z zmogljivo strojno in mrežno opremo, je postalo pravilo, da izberemo najmanjše možno zrno. Tak pristop ima namreč dve pomembni prednosti. Pri vrtanju v globino nam omogoča vpogled v najnatančnejše podatke, ki jih o določenem poslovnem subjektu beležimo, poleg tega, pa nam v primeru zahtev po novih funkcionalnosti praviloma ni potrebno spreminjati strukture področnega podatkovnega skladišča. Izkaže se, da

velikost zrna vpliva na število dimenzijskih tabel. Manjše zrno praviloma pomeni več dimenzijskih tabel in obratno [12].

Tabela dejstev, katere primer prikazujemo v preglednici 1, je sestavljena iz treh tipov atributov:

1. Zunanji ključi, ki omogočajo povezavo na dimenzijske tabele. (ZK)
2. Numerične meritve, ki jih imenujemo dejstva. (dejstvo)
3. Degenerirane dimenzije, ki ne predstavljajo meritve, a niso povezane z nobeno izmed dimenzijskih tabel. (DD)

Degenerirane dimenzije bi vsebinsko lahko zamenjali z zunanjimi ključi, a tega v praksi ne počnemo, saj bi z njim povezana dimenzija vsebovala le podatek o ključu in ničesar drugega. Iz tega razloga degenerirano dimenzijo imenujemo tudi prazna dimenzija.

Primarni ključ tabele dejstev je lahko kombinacija zunanjih ključev ali pa ena izmed degeneriranih dimenzij. Seveda to velja v primeru, ko smo prepričani v njeno enoličnost za vsak zapis tabele dejstev.

Tabela dejstev »Prodaja«
Datum (ZK)
Šifra artikla (ZK)
Šifra blagajne (ZK)
Šifra kupca (ZK)
Šifra vrste plačila (ZK)
Številka računa (DD)
Količina (dejstvo)
Vrednost (dejstvo)
Vrednost popusta (dejstvo)

Preglednica 1. Primer tabele dejstev

Za konec osnovne predstavitve tabel dejstev si pogledjmo še pojem aditivnosti. Večina uporabnih dejstev je aditivna, kar pomeni, da je rezultat seštevanja njene vrednosti po različnih dimenzijah smiseln podatek. Taka vrsta dejstev je značilna za večino transakcij, ki se beležijo v ERP sistemih združbe. Dejstva, ki jih lahko seštevamo le po določenih dimenzijah imenujemo delno aditivna. Kot primer navedemo trenutno vrednost zaloge določenega artikla, ki je aditivno po dimenziji *Artikel*, ne pa tudi po dimenziji *Datum* oziroma *Čas*. Dejstva, ki jih ni mogoče seštevati po nobeni izmed dimenzij, imenujemo neaditivna dejstva.

4.1.2. Dimenzijske tabele

Zapisi v dimenzijskih tabelah dodatno opisujejo transakcije iz tabele dejstev, s katerimi so povezani preko primarnega ključa. Dimenzijske tabele tipično vsebujejo manj zapisov od tabel dejstev, vendar pa vsebujejo več atributov, ki so največkrat opisni ali pa predstavljajo diskretne vrednosti. Njihov glavni namen je ta, da lahko pri uporabi podatkovnega skladišča preko dimenzijskih tabel omejimo oz. izluščimo zapise iz tabele dejstev, ter tako pridobimo želene informacije. Večje število atributov v dimenzijskih tabelah torej omogoča širše možnosti uporabe podatkovnega skladišča.

V nekaterih primerih bi lahko določen podatek hranili tako v tabeli dejstev, kot tudi v dimenzijski tabeli. Za odločitev, katera od možnosti je bolj primerna, se držimo naslednjega priporočila. Če atribut zavzame veliko vrednosti in je lahko osnova za nadaljnje izračune, ga hranimo v tabeli dejstev. V dimenzijsko tabelo ga uvrstimo v primerih, ko gre za atribut, ki zavzame manjše število diskretnih vrednosti ter je uporabniku lahko v pomoč pri omejevanju pogleda na zapise iz tabele dejstev.

V dimenzijskih tabelah je vrednost primarnega ključa pogosto t.i. govoreča šifra, ta pa vsebuje dodatne informacije o zapisu. V takih primerih je na podlagi take informacije smotno kreirati dodatne attribute dimenzijske tabele, ter tako bodočemu uporabniku olajšati pregledovanje podatkovnega skladišča [14]. Primer dimenzijske tabele prikazujemo v preglednici 2.

Dimenzijska tabela »Artikel«
Šifra artikla (ključ)
Opis artikla
Šifra kategorije artikla
Šifra proizvajalca artikla
Teža artikla
Priporočena cena
Število artiklov v paketu
Vrsta paketa
Datum prvega nakupa

Preglednica 2. Primer dimenzijske tabele

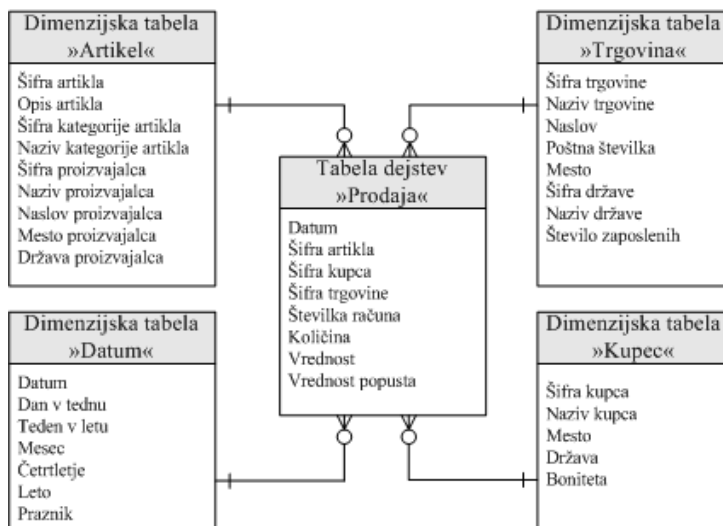
Posebna dimenzijska tabela, ki jo lahko najdemo v vsakem podatkovnem skladišču, je dimenzija *Datum*. Za razliko od ostalih dimenzijskih tabel podatkov zanjo ne pridobimo iz podatkovnih virov, ampak podatke programsko generiramo. Najpogosteje posamezen zapis v dimenziji čas predstavlja koledarski dan oz. datum, ki je hkrati tudi primarni ključ. V ostalih atributih dimenzijske tabele pa hranimo podatke kot so npr. dan v mesecu, mesec, četrtoletje, leto, dan v tednu ipd.

Preko dimenzijske tabele *Datum* smo naleteli na koncept hierarhije, ki je stalnica pri dimenzijskih tabelah, ki se nanašajo na poslovno okolje. Zaradi hierarhij se v dimenzijskih tabelah pojavljajo odvečni atributi, posledica česar so različni pristopi pri njihovem modeliranju, kar si bomo podrobneje pogledali v naslednjem razdelku.

4.1.3. Različice dimenzijskega podatkovnega modela

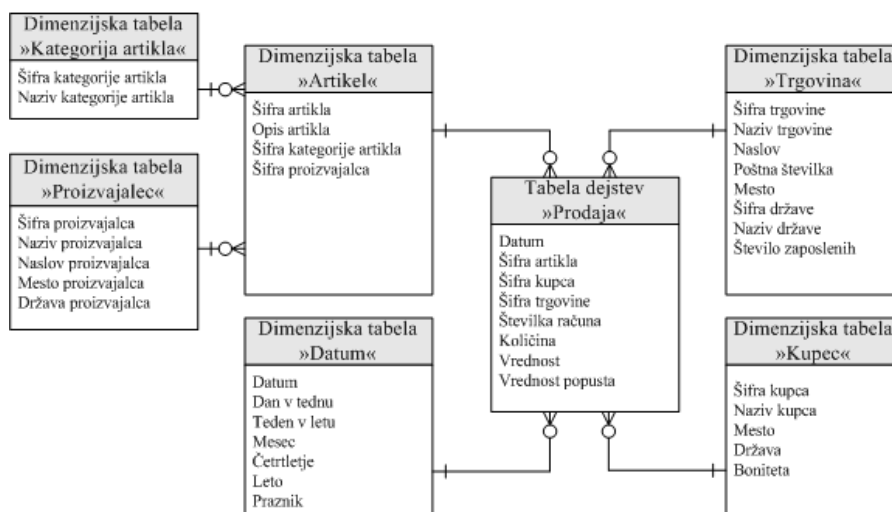
V nasprotju z relacijskim podatkovnim modelom, ki se tipično uporablja v transakcijskih sistemih, se za potrebe podatkovnega skladišča praviloma namenoma odločimo za popolno denormalizirano podatkovno strukturo, ki jo imenujemo zvezdna shema. Ta sestoji iz osrednje tabele dejstev, ki je povezana s posameznimi dimenzijskimi tabelami. Prednost uporabe zvezdne sheme je predvsem v enostavnosti podatkovnega modela.

Pričakovana prednost normalizacije dimenzijskih tabel v 3. normalno obliko, ki nas privede do podatkovnega modela imenovanega snežinkasta shema, je v zmanjšanju potrebnega podatkovnega prostora, vendar ta zaradi relativne majhnosti dimenzijskih tabel v primerjavi s tabelami dejstev ne pride do izraza. Poleg tega smo lahko v primeru sprememb kardinalnosti v podatkovnem viru pri uporabi snežinkaste podatkovne shema primorani v spreminjanje strukture podatkovnega skladišča, čemur se z uporabo zvezdne sheme elegantno izognemo [12].



Slika 3. Zvezdna shema

Pri odločanju med zvezdno shemo (slika 3) in snežinkasto shemo (slika 4) se torej nagibamo k uporabi zvezdne sheme. Če si posamezne dimenzijske tabele vseeno dovolimo delno normalizirati, pa nas to privede do hibridnega podatkovnega modela. Kot primer smiselnosti delne normalizacije navedimo dimenzijo *Artikel*, v katerem imamo med ostalimi atributi naveden tudi datum prvega nakupa. Z namenom, da uporabniku zagotovimo dodatne funkcionalnosti dimenzije *Datum*, kot je uporaba hierarhije, omenjenemu atributu ne določimo podatkovnega tipa datum, ampak vrednost atributa predstavlja ključ dimenzijske tabele *Datum*.



Slika 4. Snežinkasta shema

Za konec kot protiutež navedimo še prednost snežinkaste sheme v primerjavi z zvezdno shemo. Ker je vir za podatkovno skladišče najpogosteje visoko normaliziran transakcijski podatkovni sistem, je pri uporabi snežinkaste sheme v procesu ETL potrebno opraviti manj transformacij. Kljub navedeni prednosti pa se bomo v našem podatkovnem skladišču odločali le med uporabo zvezdne sheme in hibridnim podatkovnim modelom.

4.2. Tehnološki vidik področnih podatkovnih skladišč

Funkcionalnost ERP sistemov in sistemov za poslovno obveščanje se v določeni meri prekrivata, saj oba sistema omogočata izvajanje poizvedb, s čimer uporabniki pridobivajo informacije za odločanje. Ker sistemi poslovnega obveščanja predstavljajo specializirana orodja, namenjena izključno omenjeni nalogi, imajo pred ERP sistemom določene prednosti. Za uporabnika sta ti prednosti predvsem v večji prožnosti pri izvajanju poizvedb in bistveno hitrejši odzivnosti. Tehnologije, ki se nanašajo na področna podatkovna skladišča in omogočajo omenjeni prednosti, si bomo pogledali v nadaljevanju.

4.2.1. Tehnologija shranjevanja podatkov

Namen tehnologije OLAP (Online Analytical Processing) je shranjevanje in analiza podatkov, urejenih po pravilih dimenzijskega podatkovnega modela. Omenjeno kratico je v letu 1993 prvič uporabil Edvard Codd, kot kontrast takrat že obstoječi tehnologiji OLTP, namenjeni relacijskim podatkovnim bazam [15].

Žal pa kratica OLAP vsebinsko ne razkriva ne namena, ne bistva, ki se skriva za njeno vsebino, kar je nedvoumno podpora večdimenzionalnosti. Iz tega razloga je Nigel Pendse podal enostavno in kratko definicijo sistema OLAP, imenovano FASMI (Fast Analysis of Shared Multidimensional Information), torej kot zmožljivo večuporabniško analizo večdimenzionalnih podatkov [15].

Tehnološko se OLAP glede na način shranjevanja podatkov deli na večdimenzijski MOLAP, relacijski ROLAP in hibridni HOLAP.

MOLAP predstavlja način shranjevanja, kjer so tako podrobni kot tudi agregirani podatki shranjeni v specializirani večdimenzionalni obliki. Ker je način shranjevanja podatkov optimiziran za potrebe poizvedb v podatkovnem skladišču, gre za najhitrejšo izmed naštetih tehnologij, poleg tega pa omogoča hitro izvedbo kompleksnih izračunov. Indeksiranje agregiranih vrednosti, ki omogoča doseganje omenjene hitrosti, pa po drugi strani povzroči veliko prostorsko potratnost, kar predstavlja največjo slabost tehnologije MOLAP. Ta je še posebej izrazita v primeru velikega števila dimenzij. Podatkovno bazo tehnologije MOLAP zaradi večdimenzionalnosti imenujemo tudi OLAP kocka in predstavlja najpogosteje uporabljeno tehnologijo v podatkovnih skladiščih.

ROLAP je način shranjevanja podatkov, kjer so podrobni in agregirani podatki shranjeni v relacijski podatkovni bazi. Večdimenzijske poizvedbe, ki jih tvorijo orodja poslovnega obveščanja morajo biti pred uporabo pretvorjene v eno ali več relacijskih poizvedb. Prednost tehnologije ROLAP je v tem, da z lahkoto obvladuje zelo velika podatkovna skladišča, njena slabost pa je počasnost delovanja.

HOLAP združuje tehnologiji MOLAP in ROLAP na način, kjer so podrobni podatki shranjeni v relacijski podatkovni bazi, agregati pa v specializirani večdimenzionalni obliki.

V modernih sistemih so na voljo vse 3 izmed naštetih OLAP tehnologij. Na razvijalcu je torej odločitev, da glede na lastnosti konkretnega podatkovnega skladišča izbere najprimernejšo tehnologijo [8].

4.2.2. Agregati

Agregati, omenjeni že v prejšnjem razdelku, so izračunane vrednosti merljivih dejstev za določeno kombinacijo dimenzij. Sistem vnaprej izračunane agregate zapiše v OLAP podatkovno bazo, kar ima za posledico večjo hitrost delovanja sistema.

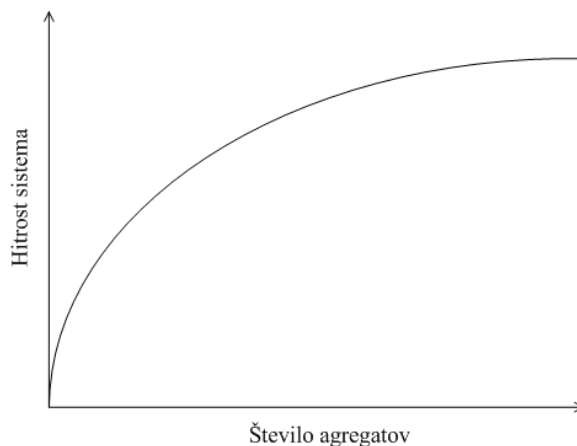
Pri implementaciji agregatov se razvijalec sooči s tehtanjem med povečanjem podatkovne baze in časom, potrebnim za preračun kocke OLAP na eni, ter hitrejšim delovanjem sistema na drugi strani. Za optimalno delovanje sistema je torej potrebno poiskati mejo, kjer je dodatno preračunavanje agregatov še smiselno. Število agregatov se s številom dimenzij ter hierarhij znotraj posameznih dimenzij hitro povečuje, kar lahko prikažemo na primeru.

V področnem podatkovnem skladišču *Prodaja* so zasnovane 3 dimenzije in sicer *Datum*, *Artikel* in *Kupec*. V vsaki izmed dimenzij je definirana hierarhija z več atributi.

- dimenzija Datum: leto, četrtoletje, mesec, dan
- dimenzija Artikel: kategorija, podkategorija, naziv
- dimenzija Kupec: kontinent, država, mesto, naziv

V tako zasnovanem sistemu lahko generiramo največ 48 agregatov ($4 \times 3 \times 4$).

Izračun števila vseh agregatov nas pri velikem številu dimenzij in hierarhij postavi v dilemo, katere agregate je smiselno tvoriti. V primeru, da izbora ne prepustimo kompleksnim algoritmom sistema, velja splošno priporočilo, da tvorimo agregate na nižjih nivojih hierarhije dimenzijskih tabel, saj na ta način pridobimo tudi na poizvedbah, ki se nanašajo na višje hierarhije. Izračun na višjih hierarhijah v tem primeru namreč poteka iz agregatov na nižjih nivojih in ne neposredno iz zrna tabele dejstev [8].



Slika 5. Prikaz povečanja hitrosti sistema s tvorjenjem agregatov

V primeru, da izbor agregatov prepustimo sistemu, lahko preračun agregatov omejimo s prostorom, ki ga za agregate dodelimo, ali pa z želenim povečanjem hitrosti sistema. Moderni sistemi omogočajo kreiranje agregatov glede na uporabo podatkovnega skladišča s strani uporabnikov, kar pomeni, da so kreirani tisti agregati, po katerih so uporabniki najpogosteje poizvedovali. Ker uporabniki ne poizvedujejo po vseh kombinacijah dimenzij enakomerno, je že z manjšim številom pravilno izbranih agregatov možno doseči opazno povečanje hitrosti sistema (slika 5).

4.2.3. Paralelizem in aktivni predpomnilnik

Paralelizem je koncept, ki omogoča povečanje hitrosti sistema na podlagi strojnih značilnosti modernih strežnikov in dimenzijskega podatkovnega modela. Z delitvijo kocke OLAP na več particij pridobimo tako na času, potrebnem za njeno polnjenje, kot tudi na odzivnosti sistema pri poizvedovanjih s strani uporabnikov.

Struktura dimenzijskega podatkovnega modela omogoča popolno paralelnost pri izračunavanju agregatov. Zaradi te lastnosti razdelitev kocke OLAP na več neodvisnih particij omogoča učinkovito izrabo večprocesorskih in več jedrnih sistemov. Za dodatno pridobitev zmogljivosti pa lahko particije razporedimo na fizično ločena diskovna polja ali v primeru velikih podatkovnih skladišč, na ločene strežnike.

Odločitev o načinu delitve oz. snovanju particij je prepuščena razvijalcu. V praksi je pogosta rešitev delitev glede na vrednost dimenzije Čas. Na ta način se v posamezni particiji nahajajo podatki za posamezno časovno obdobje, kar omogoča dodatne optimizacijske prijeme. Za vsako izmed particij lahko namreč določimo časovno periodo osveževanja kocke OLAP, kar nam omogoča, da v podatkovnih skladiščih, kjer se stari podatki ne spreminjajo osvežujemo le najnovejšo particijo. Posledica take zasnove je bistveno krajši čas, potreben za osvežitev kocke OLAP. Pravkar opisan paralelizem enako kot pri polnjenju ter osveževanju kocke OLAP omogoča tudi hitrejšo odzivnost sistema pri pregledovanju podatkov s strani uporabnikov. Slednje seveda velja le, če uporabniki istočasno izvajajo poizvedbe nad podatki, ki se nahajajo v različnih particijah.

Čeprav bomo izbor platforme za razvoj sistema poslovnega obveščanja obravnavali kasneje, si oglejmo še tehnologijo platforme *Microsoft SQL Server 2008 Analysis Services*, ki omogoča enostavno implementacijo podatkovnega skladišča v realnem času.

Ideja aktivnega predpomnilnika je v tem, da sistem ob vsaki spremembi končnih podatkov, ki so osnova za formiranje kocke OLAP, sproži avtomatsko osveževanje agregatov, ki se shranijo v aktivni predpomnilnik. Ta način, ki mu pravimo tudi avtomatski MOLAP, izniči potrebo po periodičnem osveževanju kocke OLAP po vnaprej določenem urniku, saj je za osveževanje podatkov poskrbljeno avtomatsko. V času, ko sistem pripravlja nov aktivni predpomnilnik, je v veljavi obstoječi, tako da je uporabnikom dostop do sistema omogočen neprestano, kar za klasični pristop s periodičnim osveževanjem ne velja [8].

Tehnologija aktivnega predpomnilnika je zanimiva predvsem za zahtevne uporabnike velikih sistemov, ki si v sistemu za poslovno obveščanje ne morejo privoščiti nekoliko starejših podatkov.

5. Proces ETL

Tradicionalno si vsebino procesa ETL razlagamo preko njegove kratice, ki predstavlja aktivnosti izbora, preoblikovanja in nalaganja. Po Kimbalovem pristopu, ki mu v tej nalogi sledimo [12], pa je proces ETL sestavljen iz aktivnosti izbora, čiščenja, poenotenja in dostave. Kljub temu, da tako zastavljene aktivnosti ne sovpadajo s kratico, se bomo, tako kot Kimball, vzdržali preimenovanja procesa ETL v proces ECCD.

Kot že omenjeno, Kimball ocenjuje, da za realizacijo procesa ETL v celotni realizaciji sistema za poslovno obveščanje potrošimo vsaj 70% časa, kar nedvomno priča o njegovi kompleksnosti. Smoter procesa ETL je realizacija zalednega sistema podatkovnega skladišča, ki zagotavlja vhodne podatke področnim podatkovnim skladiščem. V procesu transformacije podatkov, ki poteka od izvornih heterogenih podatkovnih virov, do končnih podatkov, urejenih v večdimenzijskem podatkovnem modelu, je pomembno zagotoviti njihovo zaščito, vsebino transformacij pa dokumentirati. Kakovostno zasnovan proces ETL v aktivnostih čiščenja in poenotenja podatkom doda informacijsko vrednost ter zagotovi eno in edino inačico resnice.

Opisane naloge, kot že rečeno, realiziramo v štirih aktivnostih:

- izbor podatkov iz heterogenih podatkovnih virov
- zagotovitev kakovosti podatkov (čiščenje)
- zagotovitev poenotenja mer in nazivov atributov preko vseh uporabljenih podatkovnih virov
- dostava podatkov v obliki večdimenzionalnega podatkovnega modela

Proces ETL lahko sprogramiramo sami, z uporabo enega izmed objektno orientiranih programskih jezikov ali pa ga realiziramo s pomočjo specializiranih orodij za podatkovno integracijo, za kar se je smiselno odločiti iz naslednjih razlogov:

- enostavnejša, hitrejša in posledično cenejša implementacija sistema
- integriran repozitorij meta podatkov, ki jih sistem ustvari avtomatsko
- učinkovitost pri izvajanju transformacij tudi nad velikimi količinami podatkov
- vgrajeno obvladovanje napak, enostavnost implementacije sprememb, vgrajen sistem razporejevalnika
- omogočanje uporabe objektno orientiranih programskih jezikov za funkcionalnosti, ki niso podprte neposredno

5.1. Shramba vmesnih podatkov

Shramba vmesnih podatkov je tesno povezana s procesom ETL, saj je namenjena shranjevanju vmesnih podatkov, ki nastajajo s transformacijami med njegovim izvajanjem. Obstajata dva pristopa pri hranjenju vmesnih podatkov. Če podatkov v shrambi med osveževanjem podatkov ne brišemo, govorimo o obstojni shrambi vmesnih podatkov. Tak pristop je smiseln v primerih, ko želimo hraniti zgodovinske podatke. V primeru, ko shrambo vmesnih podatkov med vsakim zagonom ETL procesa popolnoma prepíšemo, pa govorimo o prehodni shrambi vmesnih podatkov. Pogosto je smiseln hibriden pristop, kjer hranimo zgodovinske podatke le za določene tabele.

Pomembno je, da shrambo vmesnih podatkov ustrezno zaščitimo. Vanjo smejo dostopati le procesi ETL, ročnih vnosov podatkov ne dovoljujemo. Uporabnikom v nobenem primeru ni dovoljen dostop do shrambe, saj se v njej nahajajo vmesni podatki, katerim ni zagotovljena ne pravilnost ne konsistentnost. O spremembah strukture podatkov v shrambi končnih uporabnikov ne obveščamo.

ETL proces bi lahko zasnovali tudi brez shrambe vmesnih podatkov, vendar ima njena uporaba določene prednosti.

- V primeru napake med izvajanjem procesa ETL ni potrebno ponoviti izvajanje procesa popolnoma od začetka.
- Koristnost shrambe začasnih podatkov se izkaže pri reviziji ETL procesa ali preverjanju pravilnosti podatkov podatkovnega skladišča. S primerjanjem podatkov v različnih stopnjah transformacije je namreč mnogo lažje ugotoviti zakonitosti transformacij ali odkriti vzroke za semantične napake v podatkih.

Če se za uporabo shrambe vmesnih podatkov ne odločimo lahko nekoliko pridobimo na času, v katerem se izvede proces ETL, vendar je ta prednost pomembna le v določenih primerih.

5.2. Aktivnosti procesa ETL

V naslednjih razdelkih bomo preko aktivnosti izbora, čiščenja, poenotenja in dostave podatkov proces ETL predstavili nekoliko podrobneje.

5.2.1. Izbor podatkov

Rezultat aktivnosti izbora podatkov predstavljajo dokumenti, ki služijo kot globalni pregled nad izbranimi podatki, potrebnimi transformacijami ter strukturo podatkovnega skladišča. V aktivnosti izbora podatkov torej pripravimo okviren načrt celotnega procesa ETL in strukture podatkovnega skladišča, kar zmanjša verjetnost, da bi se bilo po začetni fizični realizaciji sistema potrebno vračati na začetek.

Aktivnost izbora podatkov delimo na opravila:

- določitev virov podatkov
- analiza kakovosti podatkov
- pregled transformacij z upoštevanjem poslovnih pravil
- pregled dimenzijskega modela
- preverba merljivih dejstev in izračunov

Aktivnost izbora podatkov se začne z določitvijo virov podatkov, za katere menimo, da bodo pripomogli k boljšem odločanju s strani uporabnikov. Podatkovni viri se lahko nahajajo v združbi ali izven nje, kar je odvisno predvsem od področja, ki ga podatkovno skladišče pokriva. V sistemih poslovnega obveščanja se večina podatkov nahaja znotraj združbe, kar pa ne pomeni, da so si v poslovnem svetu realizacije izbora podatkov podobne. Podatki se lahko namreč pojavljajo v različnih oblikah, kot so npr. tekstovne datoteke, XML datoteke, preglednice, podatkovne baze ali ERP sistemi. Specifična razporejenosti virov podatkov v

zdržbah povzroči tehnično različno zahtevne pristope pri njihovem zajemu. V primerih, ko se določen podatek v združbi nahaja na več mestih, velja pravilo, da kot vir vedno izberemo osnovni oz. prvo nastali podatek. Podatki pa se ne razlikujejo le po obliki, v kateri jih hranimo, ampak tudi po kakovosti, kar preverjamo v naslednjem opravilu aktivnosti izbora podatkov.

Kakovost podatkov razumemo kot popolnost ter vsebinsko pravilnost. Z drugimi besedami, gre za ugotavljanje anomalij na nivoju sintakse ter semantike. Za sintaktično pravilnost ne potrebujemo poglobljenih znanj o poslovanju združbe, zato je ta korak praviloma lažji. Ugotavljamo anomalije, kot so nične vrednosti atributov, v katerih bi moral biti podatek ter iščemo vrednosti, ki se ne ujemajo s pričakovanim podatkovnim tipom. Sintaktična pravilnost podatkov je zelo povezana z obliko, v kateri so podatki shranjeni. V kompleksnejših sistemih, kot je na primer ERP, so take napake bistveno redkejše. Ugotavljanje vsebinskih, oz. semantičnih napak v podatkovnem viru pa je zahtevnejša naloga, saj poleg znanj o strukturi podatkov podatkovnega vira zahteva tudi poglobljena znanja o poslovnih pravilih združbe. Med analizo, v kateri je potrebno pregledati vse vire podatkov, si za podatke v podatkovnih bazah pomagamo z obstoječimi entitetno relacijskimi diagrami. V primeru, da se v opravilu analize kakovosti podatkov izkažejo velike anomalije, moramo biti pripravljeni na možnost, da že v tej fazi prekinemo z realizacijo projekta nad temi podatki. S projektom je namreč nesmiselno nadaljevati, če kakovost podatkov ne zadošča za podporo boljšim poslovnim odločitvam.

Vsebinsko kakovosten podatek spremljamo preko vidikov:

- Pravilnosti, ki predstavlja skladnost z realnostjo.
- Nedvoumnosti, kar pomeni, da vsebine ni mogoče razumeti napačno.
- Konsistentnosti, kar pomeni, da ista informacija v zapisih nima različnih vrednosti.
- Popolnosti, kar pomeni, da za spremljane dogodke ne manjkajo vrednosti posameznih atributov ali celotni zapisi.

Ko izberemo vire podatkov in je analiza kakovosti opravljena, se osredotočimo na transformacije, ki so potrebne za preslikavo podatkov iz podatkovnega vira v večdimenzionalno podatkovno strukturo. V aktivnosti izbora podatkov seveda še ne gre za praktično kodiranje transformacij, ampak za nekoliko splošnejši pogled na transformacije podatkov, kjer pa že upoštevamo znana poslovna pravila. Anomalije v podatkih je namreč skozi proces ETL možno kvalitetno odpraviti le z upoštevanjem zapletenih poslovnih pravil, pri čemer pa je potrebno upoštevati, da je proces transformacij podatkov evolucijski proces, ki se spreminja glede na nova spoznanja o virih podatkov.

Naslednje opravilo je pregled dimenzijskega modela, ki ga na podlagi informacij o razpoložljivih podatkih in potrebah združbe okvirno določimo že pred aktivnostmi procesa ETL. Pri pregledu dimenzijskega modela je poleg informacij, pridobljenih iz predhodnih opravil nujno tudi temeljito poznavanje konceptov področnih podatkovnih skladišč, predstavljenih v četrtem poglavju te naloge.

V zadnjem opravilu s končnimi uporabniki sistema preverimo, ali so merljiva dejstva pravilno izbrana in kalkulacije pravilno dorečene. Tudi na tem mestu moramo imeti v mislih, da morajo biti končni podatki dobra podlaga za odločanje. Tak pristop nam lahko prihrani veliko časa, ki bi ga zapravili v primeru, ko bi se lotili realizacije procesa ETL z napačnimi predpostavkami.

Na podlagi omenjenih opravil procesa ETL in predhodnih aktivnosti, ki jih bomo spoznali v nadaljevanju te naloge, pripravimo dokumente, omenjene že v začetku tega razdelka. Prvi dokument, ki ga imenujemo *Matrika podatkovnega skladišča*, prikazuje seznam vseh dimenzij podatkovnega skladišča in njihovo uporabo v posameznih področnih podatkovnih skladiščih (preglednica 3). Posamezno področno podatkovno skladišče sovпада z enim ali več poslovnimi procesi v združbi.

Dimenzija / Področno podatkovno skladišče	Datum	Kupec	Dobavitelj	Artikel	Skladišče	Trgovina
Nabava	X		X	X		
Prevoz	X		X	X	X	
Prodaja	X	X		X		X

Preglednica 3. Primer matrike podatkovnega skladišča

V dokumentu *Pregled izvornih sistemov* beležimo osnovne podatke o virih, ki jih bomo uporabili pri realizaciji sistema za poslovno obveščanje:

- naziv poslovnega procesa, ki ga sistem podpira
- naziv sistema
- prioriteta oz. zaporedje pri realizaciji področnih podatkovnih skladišč
- oddelki, ki sistem uporabljajo
- ime uporabnika, ki ima znanja o poslovnih pravilih
- ime skrbnika sistema
- naziv podatkovne baze
- operacijski sistem (platforma)
- število uporabnikov
- velikost podatkovne baze

Ker so v dokumentu navedeni vsi poslovni procesi, ki jih želimo realizirati skozi več področnih podatkovnih skladišč, lahko dokument služi tudi kot okvirni pokazatelj časovnega poteka realizacije celotnega sistema.

Podrobnejše podatke, ki se nanašajo na posamezno področno podatkovno skladišče beležimo v dokumentu *Pregled toka podatkov*. V njem na nivoju posameznega atributa beležimo podatke o izvoru atributa, željeni transformaciji ter lokaciji v ponoru oz. večdimenzijskem podatkovnem modelu. V dokumentu se torej na enem mestu nahaja osnovni vpogled v zelene transformacije v procesu ETL, pri čemer pa je potrebno upoštevati, da v dokumentu niso navedene vse podrobnosti, ki so potrebne za dejansko implementacijo.

V dokumentu beležimo naslednje podatke:

- naziv izvornega sistema, kjer se nahaja atribut
- naziv izvorne tabele
- naziv atributa v izvornem sistemu
- podatkovni tip v izvornem sistemu
- opis transformacije
- naziv ciljne tabele
- naziv ciljnega atributa

- tip tabele v večdimenzijski podatkovni strukturi (dimenzija ali dejstvo)
- tip počasi se spreminjajoče dimenzije (tip 1,2 ali 3)

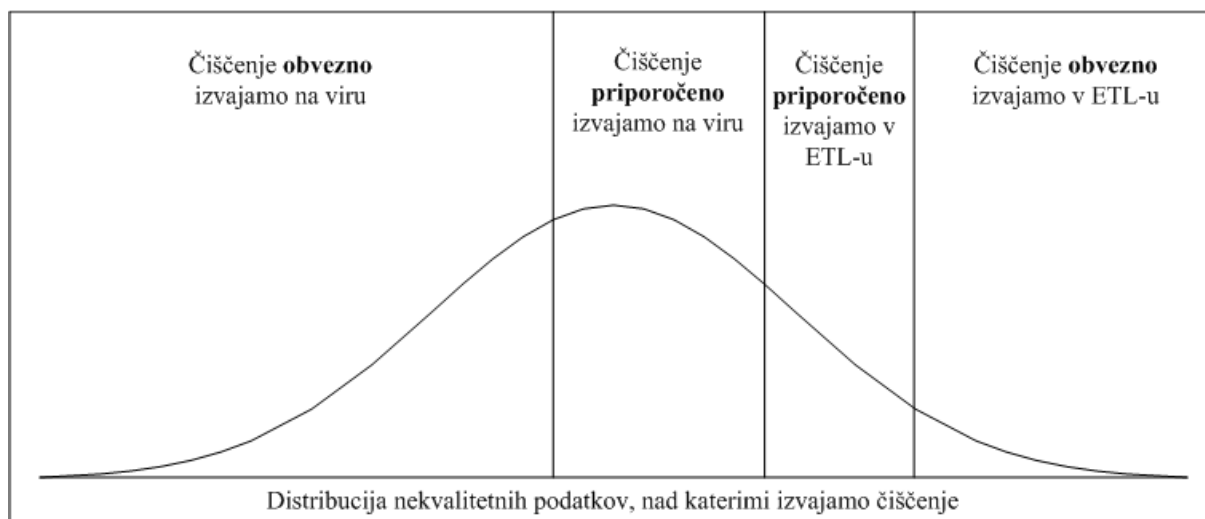
Primere dokumentov *Pregled izvornih podatkov* in *Pregled toka podatkov* si bomo pogledali v praktičnem delu naloge v nadaljevanju.

5.2.2. Čiščenje podatkov

Lastnost, po kateri se aktivnosti čiščenja in poenotenja podatkov pomembno razlikujeta od ostalih dveh aktivnosti, je vsebinsko spreminjanje podatkov. Tu torej ne gre zgolj za prenose podatkov in spreminjanje njihove strukture, ampak za spremembe, ki neposredno vplivajo na rezultate, namenjene nosilcem odločanja. Slednje priča o tem, da se je realizacije omenjenih dveh aktivnosti potrebno lotiti temeljito ter z določeno mero previdnosti.

Čiščenje podatkov je aktivnost, v kateri podatki pridobijo na vrednosti, saj skušamo izločiti nepravilnosti, ki so nastale pri njihovem zajemu. V primeru, ko podatkovno skladišče zajema zelo velike količine podatkov, je pri načrtovanju čiščenja podatkov potrebno tehtati med temeljitostjo, ter časom, v katerem se aktivnost čiščenja lahko izvede.

Pred predstavitvijo načinov čiščenja podatkov si pogledajmo mesta, kjer se čiščenje podatkov izvaja. Pri tem velja upoštevati splošno priporočilo, da se čiščenja podatkov lotimo čim bližje njihovem viru. Na sliki 6 je prikazana kategorizacija napak glede na mesto, kjer izvajamo čiščenje. Razvidno je, da je vsaj v treh četrtinah primerov napake smiselno popraviti na njihovem viru. Redki primeri, ko to ni možno, praviloma predstavljajo podatke iz zunanjih podatkovnih virov, do katerih nimamo neposrednega dostopa [12].



Slika 6. pristopi pri čiščenju podatkov

Če v podatkovnem viru odkrijemo večje nepravilnosti, ki so posledica slabe organiziranosti ali napačnega načina dela lahko ocenimo, da kakovost podatkov tega področja ne zadostuje zahtevam sistema poslovnega obveščanja. V takih primerih je možen ukrep tudi prenova poslovnega procesa.

Aktivnost čiščenja podatkov začnemo z iskanjem neakovostnih podatkov. To izvajamo tako, da med procesom ETL izvajamo teste, s katerimi odkrivamo podatke, ki jih smatramo kot neakovostne oziroma napačne. Podatke testiramo v več kategorijah glede na pravilnost vrednosti, njihove strukture in poslovnih pravil.

Pri preverjanju vrednosti iščemo napake, kot so:

- nične vrednosti v atributih, kjer to ni dovoljeno
- numerične vrednosti izven dovoljenega intervala
- vrednosti, ki jih predhodno določimo kot nedovoljene
- tekstovne napake, ki jih preverjamo s pomočjo slovarja

Preverjanje strukture se za razliko od preverjanja podatkov ne nanaša na posamezne vrednosti podatkov, ampak na njihovo povezanost. Odkrivamo napake, kot so:

- napačne ali manjkajoče povezave hierarhično urejenih podatkov
- napačne vrednosti povezanih podatkov, kot na primer mesto in poštna številka

Najzahtevnejša kategorija je gotovo odkrivanje podatkov, ki odstopajo od poslovnih pravil združbe. Teste lahko sicer realiziramo na nivoju enostavnih nepravilnosti v zapisih, kot je na primer preverba enoličnosti davčnih števil v šifrantu dobaviteljev, s poglobljenim znanjem o poslovanju združbe pa lahko zastavimo pravila, kjer kompleksnost doseže bistveno višje nivoje. Pomembno je, da prioriteto realiziramo testiranje poslovnih pravil, za katere menimo, da obstaja znatna možnost njihovih kršitev v podatkovnem viru, hkrati pa te kršitve pomembno vplivajo na končne rezultate.

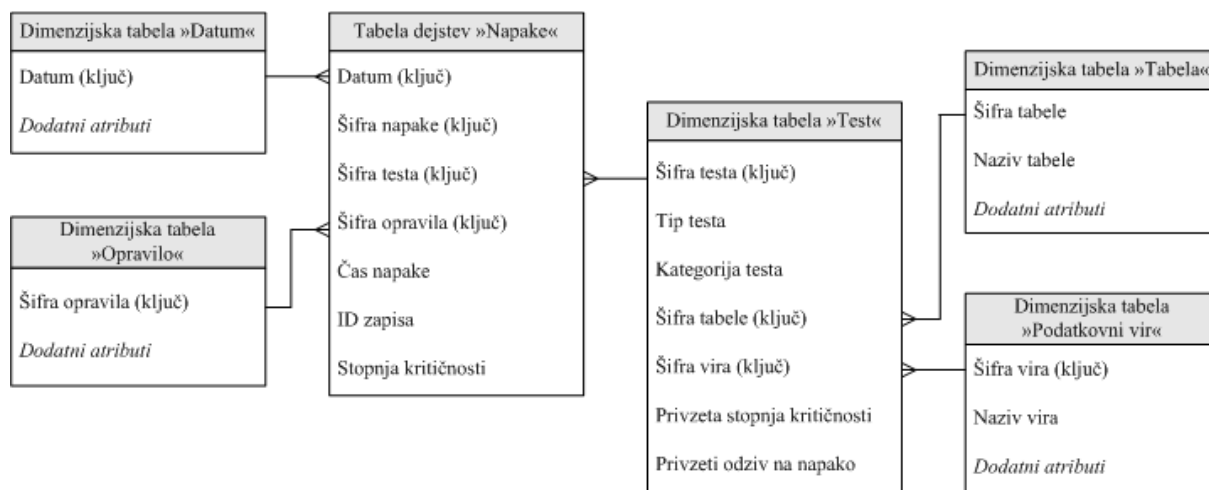
Za vsakega izmed realiziranih testov je potrebno določiti, na kakšen način se bomo med izvajanjem procesa ETL odzvali na najdeno napako. Glede na njeno kritičnost izbiramo med naslednjimi možnostmi:

- posredovanje zapisa, z dodano označbo o napaki
- zavržba zapisa
- ustavitev procesa ETL

Najpogostejša izbira je posredovanje zapisa, z dodano označbo o napaki, saj lahko na ta način z uporabo sistema poslovnega obveščanja izvedemo tudi analizo o odkritih nepravilnostih. Tak pristop nam omogoča, da v naslednjem koraku učinkovito odpravimo napake na njihovem viru. V redkih primerih, ko bi določen zapis pomembno vplival na verodostojnost podatkov v podatkovnem skladišču, celoten zapis zavržemo, kar tudi evidentiramo. V primeru kritičnih napak pa se lahko odzovemo z zaustavitvijo celotnega procesa ETL.

Vključitev metapodatkov o odkritih napakah neposredno v podatkovno skladišče je pri virih podatkov nizke stopnje kakovosti izredno pomembno, saj končnemu uporabniku omogoča pregledovanje le neoporečnih podatkov, skrbnikom sistema pa, kot že omenjeno, predstavlja orodje, ki je v pomoč pri odpravi napak na njihovem viru.

Metapodatke o napakah zapisujemo v dve podatkovni strukturi. Prvo strukturo, ki je pravzaprav ločeno področno podatkovno skladišče, v katerem se zbirajo podatki o napakah, odkritih v ETL procesu, imenujemo *Tabela napak*. Zrno *Tabele napak* predstavlja ugotovljena napaka posameznega testa. Področno podatkovno skladišče poleg tabele dejstev vsebuje še dimenzijske tabele *Datum*, *Test*, *Opravilo*, *Tabela* in *Podatkovni vir*.



Slika 6. Shema področnega podatkovnega skladišča Tabela napak

Na sliki 6 prikazano področno podatkovno skladišče *Tabela napak* vsebuje naslednje dimenzijske tabele:

- *Opravo* vsebuje podatke o posamezni izvedbi procesa ETL, kot so na primer datum in čas zagona, število prebranih zapisov ipd.
- *Tabela* identificira tabelo v procesu ETL, ki vsebuje zapis, v katerem je bila odkrita napaka.
- *Podatkovni vir* identificira podatkovni vir, iz katerega je bil črpan zapis z napako.
- *Test* vsebuje podatke o posameznem testu, ki ga izvajamo nad podatki v procesu ETL. Vsak test ima določeno stopnjo kritičnosti in odziv, ki ga sistem izvede ob identificirani napaki. Atribut *Kategorija testa* omogoča enostavnejšo izvedbo analize napak, saj omogoča analizo le nad določeno kategorijo testov, kot je npr. nekonsistentnost, nedvoumnost, pravilnost ipd.

Druga podatkovna struktura, ki jo bomo predstavili, je dimenzijska tabela *Revizija*. Vključena je neposredno v področna podatkovna skladišča in ni namenjena zgolj administratorjem sistema, kot to velja za področno podatkovno skladišče *Tabela napak*.

Dimenzija revizije, za razliko od *Tabele napak*, beleži napake le nad podatki, ki se nahajajo v tabeli dejstev posameznih podatkovnih skladišč in ne vseh tabel, ki se pojavljajo v procesu ETL. Namen dimenzijske tabele *Revizija* je evidentirati kakovost posameznega zapisa v tabeli dejstev, kar omogoča uporabnikom preiskovanje podatkov le nad podatki izbrane kakovosti.

Posamezen zapis v dimenzijski tabeli *Revizija* predstavlja določeno stopnjo kakovosti podatkov. V dimenzijski tabeli je torej toliko zapisov, kolikor različnih stopenj kakovosti podatkov v tabeli dejstev identificiramo. Število stopenj je seveda neposredno povezano s številom testov, ki jih nad podatki izvajamo.

V najenostavnejši obliki, atributi dimenzijske tabele, poleg ključa, ki dimenzijsko tabelo povezuje z zapisi v tabeli dejstev, vsebujejo le osnovne tekstualne podatke o testu. V sistemih, kjer je realizirana tudi že predstavljena *Tabela napak*, pa lahko na podlagi podatkov, ki jih iz nje pridobimo, v dimenzijsko tabelo *Revizija* shranimo končno statistiko o posameznem testu. V tabelo lahko vključimo tudi globalne metapodatke o izvedbi procesa ETL. Dimenzija revizije predstavlja tipičen primer vključitve metapodatkov neposredno v podatkovno skladišče.

5.2.3. Poenotenje podatkov

V zvezi z aktivnostjo poenotenja podatkov govorimo o dveh opravilih. Prvo predstavlja poenotenje dimenzijskih tabel, ki je izrednega pomena, saj omogoča povezovanje področnih podatkovnih skladišč. Drugo opravilo pa je poenotenje merljivih dejstev, kar z drugimi besedami pomeni, da za isto dejstvo skozi celotno podatkovno skladišče uporabljamo isti termin.

Težavnost poenotenja dimenzijskih tabel je odvisna predvsem od strukture in kakovosti podatkovnih virov. Poenotenje izvedemo tako, da iz več tabel, ki lahko pripadajo različnim sistemom, vse podatke, ki pripadajo določnemu subjektu, združimo v enoten zapis. Kot primer navedimo dimenzijsko tabelo *Kupec*, o katerem zbiramo podatke v več oddelkih združbe. Vsak oddelek zaradi lastnega interesnega področja o kupcu zbira specifične podatke. Končni rezultat poenotenja predstavlja dimenzijska tabela *Kupec*, v kateri so zbrani vsi relevantni podatki, ki zadevajo vsa področja, ki jih podatkovno skladišče pokriva. Poenotene dimenzijske tabele zagotavljajo konsistentno vsebino, ne glede na to, v katerem kontekstu oz. podatkovnem skladišču je dimenzijska tabela v uporabi.

Aktivnost poenotenja vrednosti izvedemo v treh korakih, ki jih imenujemo standardizacija, iskanje podvojenih zapisov in združevanje.

Standardizacijo vrednosti izvedemo tako, da definiramo veljavne vrednosti za vsakega izmed neskladnih atributov. Kot enostaven primer lahko navedemo oznake M, Ž, F, 0, 1, ki v različnih sistemih lahko označujejo spol subjekta. Standardizacijo izvedemo tako, da realiziramo preslikavo iz vsake najdene vrednosti v eno izmed veljavnih vrednosti. V realnih sistemih lahko pri standardizaciji naletimo na bistveno kompleksnejše in težje obvladljive nabore vrednosti, kar lahko rešujemo s specializiranimi orodji, ki za reševanje naloge uporabljajo statistične metode.

Iskanje podvojenih zapisov je podobno kot standardizacija naloga, ki se lahko izvaja na zelo različnih nivojih kompleksnosti. Primer, kjer je iskanje podvojenih zapisov izredno enostavno, so zapisi, kjer imamo o subjektu enolični identifikator, kot je na primer EMŠO ali davčna številka. V takih primerih, lahko ob zadetku z gotovostjo trdimo, da gre za podvojeni zapis.

Ker so taki primeri v realnosti redki, obstajajo za iskanje podvojenih zapisov specializirana orodja. Rezultat teh orodij je numerična vrednost, ki za posamezno kombinacijo zapisov predstavlja verjetnost, da gre za podvojeni zapis. Ta vrednost je pogosto seštevek več neodvisnih testov, ki jih orodja uporabljajo.

Združevanje zapisov predstavlja korak, ki logično sledi koraku iskanja podvojenih zapisov. Gre za postopke, v katerih določimo, katere vrednosti atributov iz podvojenih zapisov bomo v primeru nekonsistentnosti obdržali in katere zavrgli.

Zavedati se je treba, da kakovostno realizirana aktivnost poenotenja doda podatkom informacijsko vrednost, kar končnim uporabnikom sistema poslovnega obveščanja zagotavlja kakovostnejše, enovite ter nedvoumne podatke, ki so kot taki boljša podlaga za odločanje.

5.2.4. Dostava podatkov

Aktivnost dostave podatkov je zadnja aktivnost procesa ETL, v kateri podatke zapišemo v podatkovne strukture, ki smo jih predstavili v 4. poglavju. Na tem mestu velja poudariti, da se pri realizaciji velja v največji možni meri držati pravil dimenzijskega podatkovnega modela, saj nas odstopanje od ustaljenih pravil lahko prisili v uporabo ročnega programiranja. Tak pristop je v nasprotju z modernimi smernicami realizacije procesa ETL, ki temeljijo na uporabi orodij podatkovne integracije. Prednosti uporabe specializiranih orodij, ki temeljijo na intuitivnem, grafičnem vmesniku so [17]:

- avtomatska dokumentacija (metapodatki)
- enostavnejša ponovna uporaba realiziranih sklopov
- enostavnejše razumevanje ter posledično nadgrajevanje
- vgrajeno obvladovanje napak
- zagotovljena optimizacija

Kljub očitnim prednostim orodij, ki se z vsako generacijo še povečuje, pa je za zagotovitev specifičnih funkcionalnosti včasih še vedno potrebno poseči po ročnem programiranju. Pomembno je, da lahko izvajanje takih segmentov vključimo neposredno v orodja podatkovne integracije.

Pregled aktivnosti dostave podatkov začnimo z dimenzijskimi tabelami. Kljub temu, da so te po količini podatkov bistveno manjše od tabel dejstev, pa pogosto rečemo, da je podatkovno skladišče tako kakovostno, kot so kakovostne dimenzijske tabele. Za razumevanje dostave podatkov je poleg že predstavljene strukture dimenzijskih tabel potrebno poznavanje še nekaterih dodatnih konceptov.

Nadomestni ključ, ki povezuje dimenzijske tabele s tabelo dejstev, predstavlja nadomestilo za primarni ključ iz podatkovnega vira. Z razliko od primarnega ključa, ki ima lahko vsebinski pomen, je podatkovni tip nadomestnega ključa celo število brez kakršnegakoli pomena. Slednje zaradi tehničnih zakonitosti podatkovnih baz izboljša odzivnost podatkovnega skladišča. V primeru uporabe nadomestnega ključa primarni, oz. naravni ključ zapišemo v dodaten atribut dimenzijske tabele.

Če je posamezen zapis v dimenzijski tabeli statičen in se s časom ne spreminja, je relacija med naravnim in nadomestnim ključem 1 proti 1. V realnih sistemih pa se dogaja, da se dimenzije sčasoma spreminjajo, kar privede to tega, da določenemu naravnemu ključu sčasoma dodelimo več nadomestnih ključev. Ta koncept, ki predstavlja ključen razlog za uvedbo nadomestnih ključev, omogoča uporabniku, da lahko obvladuje spreminjajoče zapise v dimenzijskih tabelah. Takim tabelam pravimo počasi se spreminjajoče dimenzije.

Poznamo 3 tipe počasi se spreminjajočih dimenzij: prepisovanje (tip 1), hramba zgodovine (tip2) in nadomestna realnost (tip 3).

Tip 1 je najenostavnejša vrsta počasi se spreminjajoče dimenzije. Ko je vrednost posameznega atributa v dimenzijski tabeli potrebno dopolniti oz. popraviti in ni nobene potrebe po hrambi zgodovinskega podatka, vrednost atributa enostavno spremenimo oziroma prepišemo.

Tip 2 počasi se spreminjajoče dimenzije predstavlja razlog za uvedbo nadomestnih ključev. Izberemo ga v primerih, kjer so zgodovinske vrednosti zapisov uporabnikom pomembne. Ko se določen zapis v podatkovnem viru spremeni, v dimenzijsko tabelo dodamo nov zapis, ki vsebuje enak naravni ključ kot predhoden zapis, a nov nadomestni ključ. Tipičen primer spreminjajoče se dimenzije tipa 2 je dimenzija *Delavec*, kateremu je občasno potrebno spremeniti podatek o delovnem mestu ali naslovu stalnega prebivališča, kar predstavlja pomemben zgodovinski podatek.

Tip 3 počasi se spreminjajoče dimenzije uporabimo v primerih, ko želimo za določen podatek hraniti dve vrednosti. Kot primer navedimo dimenzijsko tabelo *Artikel*, kjer so bili artiklom dodeljene nove oznake kategorij. Če želimo ohraniti možnost, da lahko uporabnik sistema filtrira artikle tako po novih, kot tudi starih kategorijah, v obstoječi zapis dimenzijske tabele dodamo atribut, v katerega vpišemo vrednost stare kategorije. Za razliko od tipa 2, kjer za nov podatek dodamo nov zapis, gre pri tipu 3 torej za dodajanje in polnjenje novega atributa.

Na sliki 7 so počasi spreminjajo dimenzije prikazane na treh vsebinsko različnih scenarijih. Tip 1 počasi se spreminjajoče dimenzije je prikazan s spremembo številke GSM, kjer stare vrednosti ne potrebujemo. Tip 2 je ponazorjen na primeru spremembe delovnega mesta, tip 3 pa preko uvedbe novega šifranta delovnih mest.

OSNOVNI ZAPIS DIMENZIJSKE
TABELE DELAVEC

Nadomestni ključ	Naravni ključ	Ime in priimek	Številka GSM	Naziv del. mesta	Šifra del. mesta
12345	D2009010	Janez Novak	041333444	Mehanik	DM0012

TIP 1

Zaradi napačnega podatka v polju
Številka GSM, popravimo zapis.

Nadomestni ključ	Naravni ključ	Ime in priimek	Številka GSM	Naziv del. mesta	Šifra del. mesta
12345	D2009010	Janez Novak	041123456	Mehanik	DM0012

TIP 2

Zaradi spremembe delovnega mesta, v
dimenzijsko tabelo dodamo nov zapis.

Nadomestni ključ	Naravni ključ	Ime in priimek	Številka GSM	Naziv del. mesta	Šifra del. mesta
12765	D2009010	Janez Novak	041333444	Strojniki	DM0013

TIP 3

Zaradi uvedbe novega šifranta
delovnih mest, v obstoječ zapis dim.
tabele dodamo nov atribut.

Nadomestni ključ	Naravni ključ	Ime in priimek	Številka GSM	Naziv del. mesta	Šifra del. mesta	Stara šifra del. mesta
12345	D2009010	Janez Novak	041333444	Mehanik	343235	DM0012

Slika 7. Prikaz različnih tipov počasi se spreminjajoče dimenzije Delavec

Uporaba različnih tipov počasi se spreminjajočih dimenzij je povezana predvsem z naravo podatkov, njihovo strukturiranostjo v podatkovnem viru in zahtevami o dostopnosti zgodovinskih podatkov.

Pregleda dostave tabel dejstev se bomo lotili s predstavitvijo treh vrst tabel dejstev, v nadaljevanju pa bomo nato predstavili še določena praktična priporočila in pasti, na katere moramo biti pri dostavi tabel dejstev pozorni.

Najpogostejša, najbolj obširna in hkrati najnatančnejša vrsta tabel dejstev je *Transakcijska tabela dejstev*. Kot že ime pove, vsak zapis oz. zrno te vrste tabele dejstev predstavlja posamezno transakcijo. Za potrebe natančnih analiz, lahko poleg ostalih numeričnih vrednosti v transakcijski tabeli dejstev hranimo tudi natančen čas transakcije. Kot primer transakcijskih tabel dejstev lahko navedemo transakcijo prodaje, nabave, pologa, prejema, izdaje ipd.

Naslednjo vrsto tabel dejstev imenujemo *Periodični posnetek stanja*. Ta vrsta tabele dejstev je primerna za hranjenje podatkov o poslovnih področjih nad katerimi izvajamo določene periodične izračune. Gre na primer za izračunavanje kazalnikov in kazalcev, kot so na primer bonitete, ki se zaradi kompleksnosti praviloma izvajajo v transakcijskem sistemu. Izračune skupaj z ostalimi podatki, ki jih zbiramo o subjektu, v tabelo najpogosteje zapisujemo v mesečnih periodah, kar predstavlja tudi zrno tabele dejstev. Pri dostavi zapisa za tekoče obdobje imamo dve možnosti. Po enostavnejši različici zapis dostavimo šele po poteku obdobja, kar je smiselno v primerih, ko nimamo na voljo izračunov za delna obdobja. Druga možnost je dostava zapisa že pred potekom obdobja, kar pomeni, da njene vrednosti ob vsakem polnjenju podatkovna skladišča, glede na trenutno veljavne izračune, prepišemo. Merljive vrednosti tabel dejstev vrste periodični posnetek so praviloma delno aditivne, saj jih ni mogoče seštevati po dimenziji *Datum*.

Zadnjo vrsto tabel dejstev, ki je primerna za beleženje podatkov o poslovnih procesih, imenujemo *Kumulativni posnetek stanja*. Najopaznejša posebnost te vrste tabel dejstev je veliko število atributov podatkovnega tipa datum, v katerih beležimo mejnike v izvajanju posameznih aktivnosti poslovnega procesa. Posamezni zapis, ki se nanaša na konkretni poslovni proces in tako predstavlja zrno tabele dejstev, se med izvajanjem dopolnjuje z novimi vrednostmi atributov. Poleg datumskih atributov v tabeli dejstev praviloma beležimo tudi numerične vrednosti, ki se nanašajo na poslovni proces v celoti, ali pa na posamezno aktivnost. Primer poslovnega procesa, ki ga bi bilo na ta način možno obvladovati, je proces potrjevanja prejetih računov.

Na primeru lahko vse 3 vrste tabel dejstev povzamemo na naslednji način. Dvig denarja na bančnem okencu obvladujemo s transakcijsko tabelo dejstev, bonitete komitenta s periodičnim posnetkom, proces odobritve kredita pa s kumulativnim posnetkom stanja.

Pri dostavi tabel dejstev je, ne glede na njihovo vrsto, potrebno posebno pozornost nameniti referenčni integriteti, kar z drugimi besedami pomeni, da za vse tuje ključe v tabeli dejstev obstajajo pripadajoči zapisi v povezanih dimenzijskih tabelah. Kršitev integritete se lahko zgodi le ob naslednjih dveh dogodkih:

- nalaganje zapisa tabele dejstev z napačnimi tujimi ključi
- izbris zapisa dimenzijske tabele, ki je povezana z zapisom v tabeli dejstev

Ker se posledica kršenja referenčne integritete lahko kaže v napačnih končnih podatkih, ki predstavljajo osnovo za odločanje, je smiselno, da integriteto zagotovimo že na nivoju podatkovne baze. Na ta način bo vsaka kršitev sprožila napako, ter tako uporabnikom sistema poslovnega obveščanja preprečila dostop do napačnih informacij.

V primeru, ko se pri snovanju podatkovnega skladišča odločimo za uporabo nadomestnih ključev, je potrebno naravne ključe, ki jih preberemo iz podatkovnega vira, pretvoriti v nadomestne. To lahko storimo tako, da za vsak naravni ključ iz tabele dejstev v pripadajoči dimenzijski tabeli poiščemo nadomestni ključ. V primerih, ko se ta način izkaže za počasnega, kar velja za izredno obsežna podatkovna skladišča, vodimo posebno tabelo nadomestnih ključev, v kateri hranimo le naravni in nadomestni ključ. Tabelo posodobimo vsakič, ko zapišemo nov zapis v eno izmed dimenzijskih tabel ali takrat, ko se zgodi sprememba dimenzijske tabele tipa 2.

6. Orodja poslovnega obveščanja

V prejšnjem poglavju smo obravnavali proces ETL, ki je umeščen v zaledni sistem podatkovnega skladišča, do katerega končni uporabniki nimajo dostopa. V tem poglavju pa se ponovno selimo v čelni sistem, kjer bomo obravnavali orodja s katerimi končni uporabniki dostopajo do podatkov iz področnih podatkovnih skladišč. Orodja poslovnega obveščanja delimo na orodja za [6]:

- izdelavo in izvajanje poročil
- izvajanje poizvedb
- izvajanje OLAP operacij
- poslovno analiziranje

Da našeta orodja predstavljajo pomemben segment tudi za največje združbe na področju informacijske tehnologije, lahko ponazorimo z nedavnimi nakupi manjših združb, ki so razvile specializirana orodja poslovnega obveščanja kot npr. združbe *Proclarity*, ki je bila leta 2006 prevzeta s strani *Microsofta* ali združbe *Cognos*, ki jo je dobro leto kasneje ista usoda doletela s strani *IBMa* [25].

Produkt orodij za izdelavo poročil predstavljajo poročila, ki so namenjena končnim uporabnikom sistema. V preteklosti je bilo za izdelavo poročil potrebna obilica ročnega programiranja, rezultati pa so bili končnemu uporabniku dostavljeni v statični obliki. Modernjša orodja zaradi mnogo prijaznejšega uporabniškega vmesnika ne zahtevajo več poglobljenih programerskih znanj, uporabnikom pa so poročila najpogosteje na voljo kar preko spletnega vmesnika. Pri izvajanju poročil uporabniki določajo parametre, ki predstavljajo omejitve s katerimi vplivajo na vsebino poročil. S tem konceptom se je v primerjavi s predhodnimi statičnimi poročili močno zmanjšalo število zahtevanih poročil, posledično pa tudi obremenitev IT oddelka.

Orodja za izvajanje poizvedb vsebujejo grafični vmesnik, preko katerega lahko uporabniki brez poglobljenih znanj o poizvedovalnih jezikih pridobijo želene podatke. To storijo tako, da iz nabora podatkovnih elementov izberejo tiste, ki so predmet njihovega zanimanja. Podatke, ki predstavljajo rezultat poizvedbe, orodje vrne v obliki preglednice.

Orodja za izvajanje OLAP operacij so namenjena izvajanju poizvedb izključno nad večdimenzionalno podatkovno strukturo. Predstavljajo nadgradnjo orodij za izvajanje poizvedb, saj poleg izbora podatkovnih elementov, ki v tem primeru predstavljajo dimenzije in merljiva dejstva, zaradi bistveno večje odzivnosti sistema omogočajo tudi naprednejše koncepte. Med njimi velja omeniti podporo hierarhij, ki uporabniku omogočajo vrtanje v globino, vrtanje v širino med različnimi področnimi podatkovnimi skladišči ter prikaz ključnih kazalnikov poslovanja. Izredna odzivnost sistema uporabnikom omogoča hitro vrtenje in prikaz podatkov.

Orodja za poslovno analiziranje so namenjena odkrivanju vzorcev v podatkih in ustvarjanju statističnih modelov in pravil. Za ta namen se poleg statistične analize poslužujejo številnih tehnik podatkovnega rudarjenja, kot so nevronske mreže, odločitvena drevesa in linearna regresija. Novejša orodja omogočajo odkrivanje vzorcev tudi v nestrukturiranih podatkih, na primer besedilih. Modele, ki so ustvarjeni z uporabo naštetih tehnologij, lahko uporabimo za klasificiranje, pripravo priporočil za ukrepanje in napovedi dogajanj v prihodnosti. Kot

primere, kjer združbe uporabljajo orodja za poslovno analiziranje lahko navedemo iskanje goljufij pri bančnih transakcijah, napovedovanje odpovedi delovnih sredstev ali določitev storitev in blaga, ki so pisane na kožo kupcem združbe. Povprečna doba vračanja investicije za orodja poslovnega analiziranja je po raziskavah združbe IDC do petkrat krajša od ostalih orodij poslovnega obveščanja [6]. Kot slabosti orodij pa je omenjena potreba po visoko izobraženem kadru ter njihova visoka cena.

V zvezi s sistemi za poslovno obveščanje je bilo zaznati pričakovanja, da naj bi se s pomočjo naštetih, specializiranih orodij poslovnega obveščanja končni uporabniki osamosvojili, kar pomeni, da pri svojem delu ne bi več potrebovali podpore IT strokovnjakov. Žal pa so se pričakovanja vsaj v začetnih stopnjah razvoja sistemov poslovnega obveščanja izjalovila, saj so se v praksi za povprečnega uporabnika orodja izkazala kot prezahtevna.

Omenjeno težavo v veliki meri rešuje novejša generacija orodij poslovnega obveščanja, ki združujejo vsa omenjena orodja v enoten uporabniški vmesnik. Imenujemo jih nadzorne plošče. Njihova prednost je v enostavnosti uporabe in prikazu le tistih podatkov, ki jih uporabniki potrebujejo. To je omogočeno z uporabo grafičnega uporabniškega vmesnika, ki ga je mogoče popolnoma prilagoditi glede na potrebe posameznega uporabnika. Grafični vmesnik omogoča hiter in strukturiran pregled podatkov, ki so za določenega uporabnika pomembni, kar posledično pomeni hitro odkrivanje odstopanj oziroma anomalij. Nad podatki, ki jih želi uporabnik še natančneje preučiti, pa je omogočeno vrtanje v globino. Logika delovanja nadzornih plošč je zastavljena tako, da so kot orodje prilagojene načinu dela uporabnikov zato, da se uporabnikom ni treba prilagajati načinu dela orodij.

Nadzorne plošče delimo v tri kategorije [6]:

- operativne nadzorne plošče
- taktične nadzorne plošče
- strateške nadzorne plošče

Vsem naštetim kategorijam je skupno, da z različnimi poudarki omogočajo tri funkcionalnosti nadzornih plošč: nadzor, analiziranje in upravljanje. Operativne nadzorne plošče izpostavljajo funkcionalnost nadzora, taktične nadzorne plošče funkcionalnost analiziranja, strateške pa funkcionalnost upravljanja.

Operativne nadzorne plošče omogočajo nadzor tako, da na pregledni plošči prikazujejo ključne kazalnike določenega poslovnega procesa. Poleg tega se ob vnaprej določenih dogodkih prožijo opozorila, ki uporabnika opomnijo, da je potrebno ukrepati. Za operativne nadzorne plošče je pomembno, da se podatki v sistemu prenašajo čim bolj ažurno, saj je le tako omogočeno pravočasno ukrepanje. Iz tega razloga naprednejše operativne nadzorne plošče s statističnimi modeli omogočajo napovedovanje izjemnih dogodkov.

Taktične nadzorne plošče, ki so namenjene ugotavljanju uspešnosti poslovnih procesov in projektov, med vsemi tipi nadzornih plošč uporabniku ponujajo največji razpon podatkov. Sestavljene so iz dinamičnih poročil in preglednic, ki omogočajo izvajanje OLAP operacij, kar omogoča vpogled tako v visoko sintetične, kot tudi najbolj podrobne podatke. Naprednejše taktične nadzorne plošče omogočajo modeliranje s t.i. kaj-če scenariji, kjer sistem vrne vrednosti določenih kazalnikov glede na hipotetično spremenjene vhodne parametre.

Strateške nadzorne plošče, ki veljajo za najpopularnejše izmed vseh tipov nadzornih plošč, tipično prikazujejo ključne kazalnike, ki zajemajo poslovanje celotne združbe ali posameznih oddelkov. Kazalniki so na pregleden način primerjani z načrtovanimi vrednostmi. Strateške nadzorne plošče pogosto vsebujejo tudi standardna obdobja poročila, saj se ta izkažejo kot najprimernejša za ravnatelje na najvišjih nivojih združbe. Ključnega pomena za uspešno implementacijo je, da strateške nadzorne plošče omogočajo odkrivanje vzrokov za težave, ki jih prikažejo odstopanja od predvidenih vrednosti kazalnikov.

6.1. Uporabniki orodij poslovnega obveščanja

Naštete kategorije orodij poslovnega obveščanja omogočajo različne funkcionalnosti, ki so namenjene različnim vrstam uporabnikov. Iz tega razloga se je kot zmotno izkazalo prepričanje, da lahko z eno vrsto orodij zadovoljimo vsakega izmed njih. Kot že rečeno, se kot najbolj univerzalno orodje izkažejo nadzorne plošče. Te uporabnikom z različnimi potrebami omogočajo predstavitev tistih podatkov, ki jih dejansko potrebujejo ter pregledovanje podatkov na načine, ki jih lahko samostojno obvladajo.

Na sliki 8 prikazujemo delitev uporabnikov orodij na ustvarjalce podatkov in njihove potrošnike. Med ustvarjalce podatkov uvrščamo uporabnike orodij za poslovno analiziranje in orodij za izdelavo poročil. To so najpogostejše poslovni analitiki, ki so se priučili uporabe zahtevnejših orodij in pri nadaljnji obdelavi podatkov pogosto uporabljajo preglednice ali strokovnjaki s področja informacijske tehnologije. Medtem, ko slednje uvrščamo izključno v skupino ustvarjalcev podatkov, pa so poslovni analitiki praviloma tako ustvarjalci, kot tudi potrošniki podatkov.



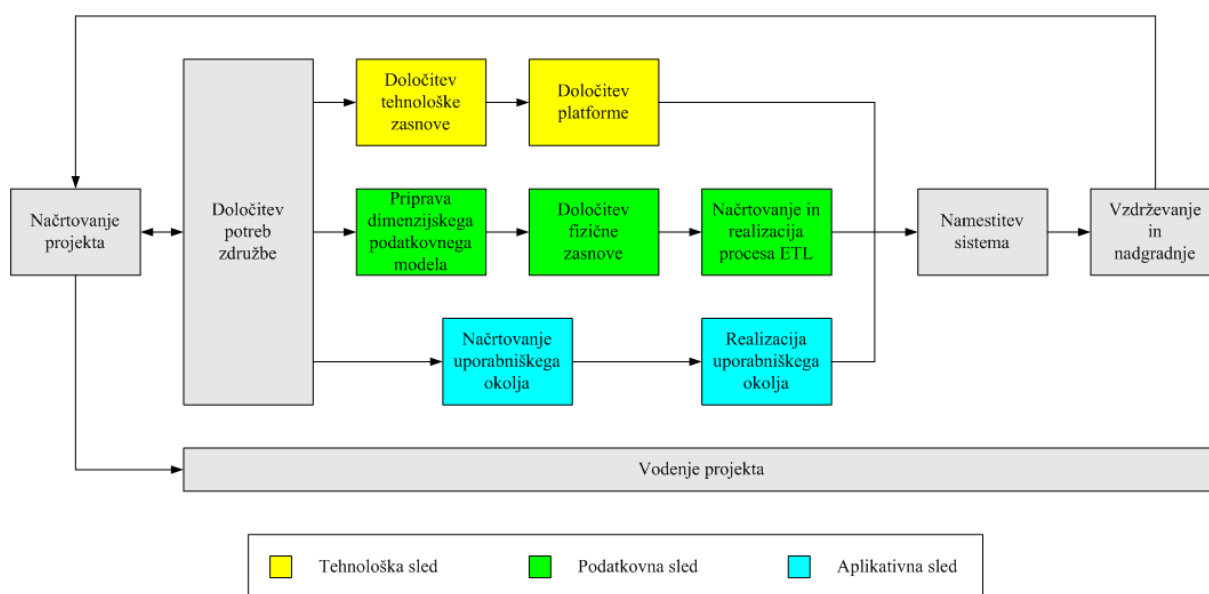
Slika 8. Delitev uporabnikov glede na uporabo orodij poslovnega obveščanja

Skupina potrošnikov podatkov, ki predstavlja 80% vseh uporabnikov, za svoje delo uporablja orodja za izvajanje poročil in poizvedb ter OLAP operacij, kar je, kot že omenjeno, lahko učinkovito združeno v nadzornih ploščah. Tipični porabniki podatkov so ravnatelji združbe na različnih nivojih ter ostali uporabniki sistema poslovnega obveščanja znotraj združbe in izven nje [6].

7. Pristop življenjskega cikla

Do sedaj smo v naši nalogi obravnavali sistem za poslovno odločanje iz vidika komponent, iz katerih sestoji. V prihajajočem poglavju pa se bomo posvetili metodologiji razvoja sistema, konkretnije pristopu življenjskega cikla [13], katere avtor je že nekajkrat omenjeni Ralph Kimball. Metodologija določa niz med seboj povezanih aktivnosti, ki jih je potrebno izvesti za uspešno realizacijo sistema poslovnega obveščanja.

Aktivnosti, ki jih prikazujemo na sliki 9, se lahko izvajajo sočasno. Razdeljene so na tehnološko, podatkovno in aplikativno sled. Razumeti je potrebno, da metodologija ne določa natančnega terminskega plana za izvedbo aktivnosti, saj je to odvisno od značilnosti posameznega projekta. Kot pri vsaki metodologiji je tudi tu potrebna določena mera kreativnosti, ki metodologijo približa potrebam konkretnega projekta.



Slika 9. Aktivnosti metodologije življenjskega cikla

Poudariti velja, da Kimballova metodologija življenjskega cikla predvideva iterativno realizacijo podatkovnega skladišča. Postopna realizacija manjših, obvladljivih področnih podatkovnih skladišč ima torej prednost pred enotnim, velikim projektom, kjer bi bilo celotno podatkovno skladišče nameščeno naenkrat. Realizacijo sistema poslovnega obveščanja lahko torej razumemo kot proces in ne običajen projekt s predvidenim začetkom in koncem.

Metodologija življenjskega cikla je zasnovana za podporo razvoja tudi največjih projektov sistema poslovnega obveščanja. Iz tega razloga je seveda dopuščeno in celo priporočljivo, da se pri konkretnih, manjših projektih, metodologija priredi dejanskim potrebam.

7.1. Načrtovanje in vodenje projekta

Aktivnost načrtovanja projekta se začne z ugotavljanjem potreb združbe po funkcionalnostih, ki jih nudi sistem za poslovno obveščanje. Pri poslovnih uporabnikih lahko naletimo na naslednje odzive:

- poslovni uporabniki ne čutijo potrebe po sistemu
- le posamezni vodje oddelkov so zainteresirani za sistem
- večina vodij oddelkov zahteva vzpostavitev sistema

Ne glede na odziv poslovnih uporabnikov je pred začetkom projekta smiselno preveriti tudi zrelost združbe, kar storimo s preverjanjem petih pogojev:

- projekt uživa podporo vodstva združbe
- cilji sistema poslovnega obveščanja sovpadajo s strategijo združbe
- oddelek za informatiko in poslovni uporabniki dobro sodelujejo
- poslovni uporabniki zahtevajo boljše informacije za odločanje
- izvorni podatki so dovolj kakovostni za uspešno realizacijo sistema

V primeru, ko so pogoji zadovoljivo izpolnjeni, lahko z načrtovanjem projekta nadaljujemo, v nasprotnem primeru pa je smiselno, da že na tej točki razmislimo o preložitvi oziroma ustavitvi projekta. Na ta način se namreč izognemo nepotrebnim stroškom, ki bi nastali, če bi bili primorani projekt ustaviti kasneje.

Načrtovanje projekta nadaljujemo z opraviлом okvirne določitve obsega projekta, ki poleg okvirnih časovnih rokov zajema še predvidene funkcionalnosti. Ker gre pri razvoju sistema za poslovno obveščanje, kot že omenjeno, za iterativni razvoj, lahko že v začetku določimo predvidene obsege za več bodočih iteracij. V primerih, ko za potrditev s strani vodstva združbe potrebujemo dodatno motivacijo, je na tem mestu smiselno pripraviti izračun časa povrnitve investicije.

Za izdelavo načrta projekta je potrebno izvesti naslednja opravila:

- izbrati naziv projekta
- določiti projektno skupino s pripadajočimi vlogami
- izdelati terminski plan

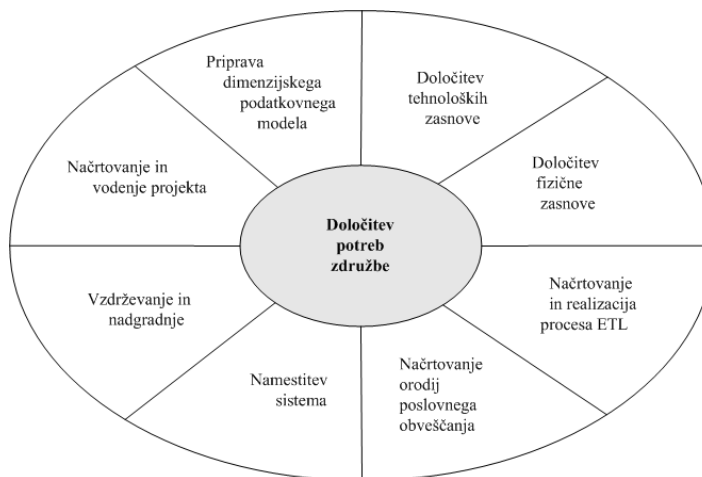
Pri izdelavi terminskega plana se priporoča, da tega izdelamo za čim več iteracij življenjskega cikla. Pri takem pristopu pa se je potrebno zavedati, da gre pri kasnejših iteracijah zgolj za ocenjene aktivnosti, kar je na terminskem planu potrebno tudi ustrezno označiti.

Smoter aktivnosti vodenja projekta je v tem, da s sprotim preverjanjem stanja projekta skušamo zadostiti ciljem projekta v čim večji možni meri. Napredovanje projekta se preverja na rednih sestankih projektne skupine, kjer se beleži napredek in morebitna odstopanja od načrta. Glede na odstopanja, nove ugotovitve in prihodnja pričakovanja je potrebno popravljati predviden obseg in terminski plan projekta.

Projekt vzpostavitve sistema za poslovno obveščanje ima določene posebnosti, med katerimi lahko izpostavimo predvsem potrebo po visoki povezanosti poslovnih uporabnikov in razvijalcev sistema, iterativno naravo projekta ter pogoste nepredvidene težave z izvornimi podatki, ki lahko bistveno podaljšajo trajanje projekta.

7.2. Določitev potreb združbe

Potrebe združbe, ki jih razumemo kot potrebe poslovnih uporabnikov, vplivajo na večino odločitev pri realizaciji sistema poslovnega obveščanja. Potrebe uporabnikov neposredno vplivajo na podatke oz. področja, ki naj jih podatkovno skladišče podpira, pogostost osveževanja podatkovnega skladišča ter specifikacijo v zvezi z orodji poslovnega obveščanja. Število uporabnikov vpliva na strojne zahteve in določitev platforme, naknadne potrebe ali zahteve uporabnikov pa vplivajo na vzdrževanje in nadgradnjo sistema. Iz povedanega lahko sklepamo, da potrebe poslovnih uporabnikov posredno ali neposredno vplivajo na vse vidike projekta realizacije sistema poslovnega obveščanja, kar je prikazano na sliki 10.



Slika 10. Določitev potreb združbe v središču projekta

Potrebe poslovnih uporabnikov lahko pridobimo izključno s komunikacijo, pri čemer pa lahko naletimo na določene težave. Te so še posebej izrazite v primerih, ko je prepad med znanjem IT strokovnjakov in poslovnih uporabnikov preglobok. Nerazumevanje poslovnih procesov s strani IT strokovnjakov in hkratno nerazumevanje osnovnih principov ter posledično zmožnosti, ki jih ponuja sistem poslovnega obveščanja s strani poslovnih uporabnikov, lahko resno ogrozi uspešno določitev potreb združbe in posledično uspešnost celotnega projekta.

Poznamo dve tehniki za določitev potreb poslovnih uporabnikov in sicer na podlagi intervjujev ali vodenih sestankov. Prednost prvega pristopa je v manjši obremenitvi poslovnih uporabnikov, saj ti za individualne intervjuje potrošijo manj časa, kot bi ga v primeru sestankov, na katerih je prisotnih več udeležencev. Kot dodatno prednost pristopa z intervjuji pa velja omeniti še večji obseg pridobljenih povratnih informacij, ki jo dosežemo zaradi posamične obravnave poslovnih uporabnikov. Poglavitna prednost drugega pristopa, torej pristopa z vodenimi sestanki je v tem, da zaradi množične udeležbe spodbuja kreativno razmišljanje, kar lahko vodi k učinkovitim rešitvam. Poudariti velja, da je izbira primerne pristopa odvisna predvsem od izkušenj ljudi in značilnosti združbe.

Informacije, zbrane na podlagi intervjujev ali sestankov, zberemo v dokumentu, v katerem zabeležimo ugotovljene potrebe poslovnih uporabnikov. Tak pristop nam omogoča, da imamo na enem mestu zbrane vse potrebe združbe, poleg tega pa ti zapisi omogočajo dodatno preverbo, ali so bile potrebe poslovnih uporabnikov pravilno razumljene in evidentirane.

7.3. Tehnološka sled

Tehnološka sled sestoji iz opravila določitve tehnološke zasnove in opravila določitve platforme, na kateri se bo izvajal sistem za poslovno obveščanje.

Na določitev tehnološke zasnove vplivajo trije dejavniki:

- potrebe poslovnih uporabnikov
- obstoječa IT infrastruktura
- strategija združbe za področje IT tehnologije

V aktivnosti določitve tehnološke zasnove je potrebno pripraviti dokumenta *Načrt tehnološke arhitekture* in *Načrt infrastrukture*. *Načrt tehnološke arhitekture* je dokument, v kateri določimo komponente sistema in njihove povezave ter pripadajoče funkcionalnosti. Pri določitvi komponent na tem mestu še ne gre za konkretne produkte, ampak umestitev komponent za shranjevanje podatkov in storitev, ki podpirajo določeno funkcionalnost. Komponente uvrstimo v čelni ali zaledni sistem poslovnega obveščanja. Načrt tehnološke arhitekture lahko poleg omenjenega vsebuje tudi povzetek poslovnih potreb, poslovna pravila in določitev metapodatkov.

Priprava načrta tehnološke arhitekture sestoji iz naslednjih opravil:

- določitev ekipe za pripravo načrta tehnološke arhitekture
- zagotovitev informacij o infrastrukturi in strategiji združbe
- določitev kritičnih dejavnikov in njihova preverba s poslovnimi uporabniki
- priprava osnovne sheme tehnološke arhitekture
- določitev korakov za realizacijo posameznih komponent načrta

V načrtu infrastrukture, ki je lahko zajet v načrtu tehnološke arhitekture, je potrebno določiti strojno opremo ter pripadajoče operacijske sisteme, mrežno opremo in osnovne karakteristike računalnikov odjemalcev. Načrt infrastrukture in tehnološke arhitekture morata biti med seboj usklajena, saj določitev infrastrukture neposredno vpliva na tehnološko arhitekturo.

Naslednje opravilo tehnološke sledi je določitev platforme, oziroma konkretnih produktov, ki morajo zadovoljiti pripravljene načrte tako s stališča poslovnih, kot tudi tehnoloških potreb.

Produkte je potrebno izbrati za naslednja področja sistema poslovnega obveščanja:

- sistem za upravljanje podatkovnih baz
- podatkovna integracija (ETL)
- orodja poslovnega obveščanja

Izbor produktov se izvede tako, da se odločitev pripravi na podlagi raziskave tržišča in odločitvenih kriterijev, ki jih določimo za vsakega izmed njih. V primeru obsežnega projekta lahko za posamezen produkt pripravimo tudi prototip sistema, kar nam olajša izbiro, vendar pa lahko občutno podaljša trajanje izbora in posledično trajanje celotnega projekta.

V praksi se izkaže, da se izbor le redko izvaja na tako sistematičen način, saj nanj pogosto pomembno vplivajo dejavniki, kot so obstoječa strojna in programska oprema, prednostno obravnavanje določenih produktov oz. proizvajalcev in obstoječa znanja in izkušnje projektne ekipe z določenim produktom.

7.4. Podatkovna sled

Preden se poglobimo v posamezne aktivnosti podatkovne sledi, si pogledjmo še prednosti in pomanjkljivosti dveh možnih pristopov k realizaciji sistema. Pristop, ki ga zagovarja Kimball, imenujemo pristop od spodaj navzgor in predstavlja pristop od izvajalnega k strateškemu nivoju, Inmon pa nasprotno zagovarja pristop od zgoraj navzdol, torej pristop od strateškega k izvajalnemu nivoju [23].

Pristop od spodaj navzgor lahko slikovito predstavimo z reko »Z majhnimi koraki do visokih ciljev«. Po Kimballovem pristopu se namreč projekt tipično začne na podlagi iniciative določenega oddelka v združbi, kjer se na podlagi želje po kvalitetnejših informacijah za odločanje pristopi k realizaciji področnih podatkovnih skladišč. Ker tako zasnovan projekt praviloma ni obsežen, je lahko ta zaključen hitro ter z relativno nizkimi stroški. V primeru, ko je začetni projekt uspešen, obstaja velika verjetnost po širitvi sistema, saj bodo zahteve po širitvi sistema na ostala področja delovanja združbe hitro prišla na dan.

Kot največjo pomanjkljivost Kimballovega pristopa lahko izpostavimo bojazen, da področna podatkovna skladišča ostanejo med seboj nepovezana. Iz tega razloga je potrebno že pri snovanju prvih področnih podatkovnih skladišč imeti v vidu nadaljnje možne širitve sistema. Sposobnost predvidevanja bodočih potreb in zahtev predstavlja največji izziv pristopa od spodaj navzgor. Projektna ekipa mora namreč že od samega začetka zagotavljati konsistentnost podatkov in skladnost skupnih dimenzijskih tabel, saj lahko le tako zagotovimo učinkovito povezavo področnih podatkovnih skladišč v uporabniku prijazen ter enovit sistem poslovnega obveščanja.

Pristop od spodaj navzgor je torej privlačen, saj zagotavlja hitre rezultate z nizkim finančnimi tveganji. Težavo pa, kot že rečeno, predstavlja zagotavljanje učinkovite povezanosti ločeno zasnovanih področnih podatkovnih skladišč.

Inmonov pristop od zgoraj navzdol, za razliko od Kimballovega, predvideva namestitve celotnega sistema v enem koraku, kar je moč doseči le s temeljitim načrtovanjem. Sistem sloni na realizaciji normalizirane podatkovne strukture, imenovane *Podatkovno skladišče združbe*. Ta predstavlja izvorni sistem področnim podatkovnim skladiščem, v katerih so podatki, enako kot pri Kimballovem pristopu, urejeni v zvezdni ali snežinkasti shemi.

Pri Inmonovem pristopu torej ni nikakršne bojazni, da področna podatkovna skladišča ne bi bila med seboj povezana in usklajena, saj imajo enoten podatkovni vir. Veliko težavo pa lahko predstavlja obsežnost projekta, saj pri tem pristopu verjetnost, da projekt ne bo nikoli zaključen, ni zanemarljiva. Ker pred samim zaključkom projekta nimamo nobenih povratnih informacij s strani uporabnikov sistema, se nam lahko pripeti, da tudi v primeru uspešno zaključenega projekta ne bomo zadostili pričakovanj uporabnikov.

Poudariti je potrebno, da v praksi le redko sledimo izključno enemu izmed omenjenih pristopov. Smotrno je namreč iz vsakega prevzeti najboljše, pri čemer upoštevamo posebnosti združbe in konkretnega projekta. Za potrebe naše naloge bomo, kot že rečeno, v veliki meri sledili Kimballovem pristopu.

7.4.1. Priprava dimenzijskega podatkovnega modela

Znanja o osnovah dimenzijskega modela smo pridobili v podpoglavju 4.1. V tem razdelku bomo predstavili postopke, ki omogočajo učinkovito pripravo konkretnih dimenzijskih modelov in pripadajočih dokumentov. Pomembno je razumeti, da je priprava dokumentov v celotni podatkovni sledi specifičen proces, v katerem dokumente zaradi novih spoznanj pogosto obnavljamo in dopolnjujemo. Ni torej izjema, da se z namenom uskladitve dokumentov z želenimi spremembami, na primer s podporo novim funkcionalnostim, v podatkovni sledi vračamo na predhodne aktivnosti.

Ob predpostavki o temeljitem poznavanju potreb združbe in obstoječih podatkovnih virih se lotimo priprave *Matrike podatkovnega skladišča*, ki smo jo predstavili v podpoglavju 5.2. *Matrika podatkovnega skladišča* predstavlja dobro izhodišče za vse sledeče dokumente, saj omogoča jasen pogled na vsa njegova področja in notranje povezanosti. Pripravo *Matrike podatkovnega skladišča* začnemo s popisom poslovnih področji.

Priporoča se, da tako pri popisu kot pri realizaciji področnih podatkovnih skladišč začnemo s tistimi, ki imajo za osnovo le en podatkovni vir. Na ta način zmanjšamo verjetnost, da bi zaradi prevelike kompleksnosti projekt zastal že na samem začetku. Pripravo nadaljujemo s popisom dimenzij, ki jih nato v zadnjem koraku povežemo s področnimi podatkovnimi skladišči in tako tvorimo matriko.

Matrika podatkovnega skladišča nam nudi pomembne informacije o gostoti oziroma stopnji povezanosti posameznih področnih podatkovnih skladišč, poleg tega pa nam jasno razkriva, katere izmed dimenzij so uporabljene v mnogih področnih podatkovnih skladiščih. Pri takih dimenzijah je potrebno biti še posebej pazljiv, da zagotovimo njihovo medsebojno skladnost.

Ko so področna podatkovna skladišča in pripadajoče dimenzije določene se lahko lotimo natančnejšega načrtovanja posameznih zvezdnih shem.

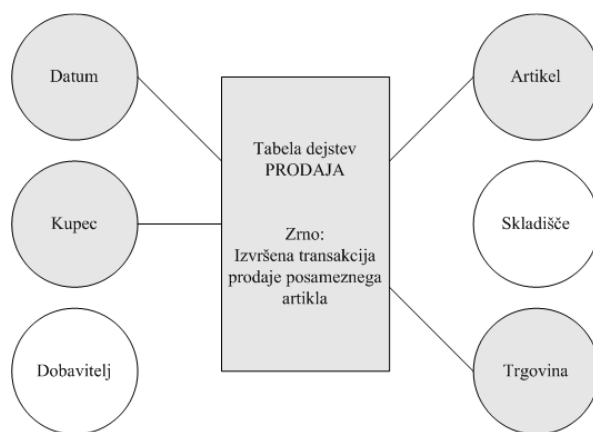
Zvezdne sheme načrtujemo v štirih korakih:

- izbor poslovnega procesa
- določitev zrna tabele dejstev
- določitev dimenzij glede na izbrano zrno
- določitev merljivih dejstev

Ker določitev zrna tabele dejstev neposredno vpliva na dimenzije, se lahko zgodi, da je potrebno ob izboru zrna popraviti matriko podatkovnega skladišča.

V povezavi s tabelami dejstev metodologija življenjskega cikla predvideva izdelavo *Diagrama tabele dejstev* (slika 11) in *Podrobnega diagrama tabel dejstev* (slika 12).

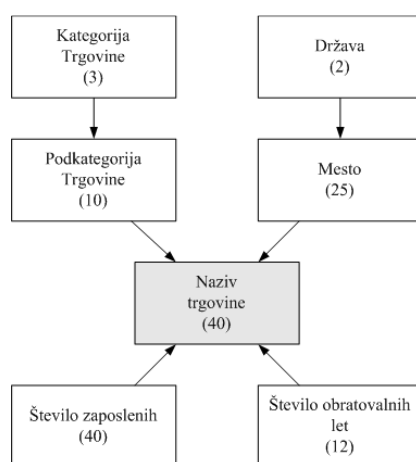
V prvem dokumentu zabeležimo zgolj osnovne podatke o tabeli dejstev, kot so naziv tabele, opis tabele, opis zrna ter prikažemo povezave na dimenzijske tabele, s katerimi je tabela dejstev povezana. V diagramu navedemo tudi dimenzijske tabele, ki s tabelo dejstev niso povezane.



Slika 11. Primer diagrama tabele dejstev Artikel

V *Podrobnem diagramu tabele dejstev* poleg naziva navedemo tudi vse attribute tabele dejstev. Vsak atribut uvrstimo v enega izmed treh tipov atributov: ključ dimenzijske tabele, osnovno dejstvo ali izpeljano dejstvo. Razlika med osnovnimi in izpeljanimi dejstvi je v tem, da prva pridobimo neposredno iz podatkovnega vira, za izračun vrednosti izpeljanih dejstev pa je potrebno opraviti določene izračune ali transformacije. Izpeljana dejstva delimo na aditivna dejstva, katerih vrednost lahko izračunamo na podlagi vrednosti ostalih dejstev posameznega zapisa in neaditivna, ki za izračun praviloma zahtevajo kompleksnejše prijeme. Kot primer neaditivnih izpeljanih dejstev lahko navedemo najrazličnejše kazalnike in kazalce. Pri načrtovanju lahko izpeljanim dejstvom namenimo posebno pozornost, tako da v posebni preglednici posameznemu izmed njih določimo še način izračuna, aditivnost, algoritem za izračun in potrebne transformacije za izračun.

Pri načrtovanju dimenzijskih tabel je potrebno pripraviti *Podrobni diagram dimenzijskih tabel* (slika 12). Vsako dimenzijo predstavimo na lastnem diagramu, kjer je poleg posameznih atributov dimenzijske tabele razvidna tudi njihova hierarhična odvisnost in ocena različnih vrednosti določenega atributa. Podobno kot pri tabelah dejstev lahko tudi za dimenzijske tabele pripravimo dodatno preglednico, kjer za vsak atribut beležimo podatke, kot so: naziv atributa, definicija atributa, primer podatkov, tip počasi se spreminjajoče dimenzije.



Slika 12. Primer podrobnega diagrama dimenzijske tabele Trgovina

Pomembno je, da vse pripravljene dokumente predstavimo tudi predstavnikom poslovnih uporabnikov. Preko vodenih sestankov lahko ugotovimo določene pomanjkljivosti načrta ali želje po dodatnih funkcionalnostih, ki jih upoštevamo pri popravkih načrta.

7.4.2. Določitev fizične zasnove

Aktivnost določitve fizične zasnove izvedemo preko petih opravil:

- določitev standardov za imenovanje objektov
- načrtovanje fizičnega modela
- načrtovanje agregatov in indeksov
- določitev tehničnih lastnosti podatkovne baze
- določitev orodij za spremljanje uporabe

Pri določitvi standardov se osredotočamo predvsem na imenovanja tabel in atributov. Pri tem moramo tehtati med dolgimi podrobnimi nazivi, ki so za uporabnike lahko nepregledni ter kratkimi nazivi, kjer pa je vprašljiva njihova pravilna interpretacija. Določitev standardov za imenovanje je pomembna predvsem zaradi poenotenja imen tako v shrambi vmesnih podatkov, kot tudi v končnih tabelah, pripravljenih za dostavo.

Fizični podatkovni model se od logičnega razlikuje v tem, da se za posamezne attribute določijo dodatne lastnosti. Pri tem je poleg določitve primarnega ključa in dovoljenja za vnos ničelnih vrednosti pomembna predvsem določitev podatkovnega tipa. Pri snovanju fizičnega podatkovnega modela je smiselno uporabiti orodja za podatkovno modeliranje, ki so sestavni del sistemov za upravljanje podatkovnih baz. S tako integracijo namreč zagotovimo konsistentnost modela v vseh fazah razvoja.

Agregate smo podrobneje opisali v razdelku 4.2.2. Na tem mestu torej velja le povzeti, da gre za koncept, s katerim preko vnaprejšnjih izračunov povečamo hitrost delovanja sistema. Podobno velja za načrtovanje indeksov v relacijskih podatkovnih bazah. Pohitritve se v tem primeru največkrat nanašajo na transformiranje podatkov med samim procesom ETL, v primeru uporabe tehnologije ROLAP pa tudi na hitrost izvajanja področnih podatkovnih skladišč. Ker je uporaba tehnologije ROLAP prej izjema kot pravilo, končni uporabniki pohitritev na podlagi določitve indeksov ne občutijo, saj se podatki v področnih podatkovnih skladiščih praviloma ne nahajajo v relacijski podatkovni bazi, ampak v oblikah MOLAP oz. HOLAP, opisanih v razdelku 4.2.1.

K tehničnim lastnostim podatkovne baze prištevamo lastnosti, kot so velikosti indeksov, delitev podatkovne baze na particije, določitev RAID polj, nastavitve arhiviranja, nastavitve dnevnika podatkovne baze in podobno. Ker gre za tehnične koncepte, ki z vsebino sistema poslovnega obveščanja niso neposredno povezani, se v njihove podrobnosti ne bomo spuščali.

Določitev orodij za spremljanje uporabe sistema za poslovno obveščanje s strani uporabnikov je pomembna iz več vidikov. Omogoča nam optimizacijo nastavitvev preko že omenjenega koncepta agregatov, saj lahko povečamo hitrost izvajanja tistih poizvedb, ki jih uporabniki najbolj pogosto izvajajo. Spremljanje uporabe omogoča tudi boljšo podporo uporabnikom, saj se lahko osredotočimo na področja, do katerih najpogosteje dostopajo. Kot dodaten razlog za zgodnjo realizacijo spremljanja uporabe sistema s strani uporabnikov, navedimo še statistiko

uporabe, ki je lahko dobra osnova za naknadno upravičenje projekta sistema poslovnega obveščanja pri vodstvu združbe.

7.4.3. Načrtovanje in realizacija procesa ETL

Realizacijo procesa ETL smo preko opravil izbora, čiščenja, poenotenja in dostave podatkov podrobno obravnavali v podpoglavju 5.2. Iz tega razloga se bomo na tem mestu osredotočili na opravila, ki jih izvedemo v okviru načrtovanja procesa ETL:

- izdelava grobega načrta procesa ETL
- končni izbor ter testiranje orodja za podatkovno integracijo
- priprava podrobnega načrta procesa ETL

Prvo opravilo načrtovanja je izdelava grobega načrta procesa ETL, ki na diagramu prikazuje povezavo med tabelami v podatkovnih virih in dimenzijskih tabelah ter tabelah dejstev v področnih podatkovnih skladiščih. Gre pravzaprav za nekoliko poenostavljen grafični prikaz podatkov, ki jih zberemo v dokumentu *Pregled toka podatkov*, obravnavanemu v razdelku 5.2.1.

V aktivnosti načrtovanja procesa ETL imamo že dovolj informacij o sistemu, da lahko v okviru končnega izbora ter testiranja orodja za podatkovno integracijo ponudnikom orodij zastavimo konkretna vprašanja o prednostih in slabostih orodij, ki jih ponujajo. V primeru večjih sistemov se lahko odločimo za realizacijo prototipov podatkovne integracije z različnimi orodji, s čimer dobimo še jasnejšo sliko njihovih prednosti oziroma pomanjklivosti.

Zadnje opravilo v okviru načrta procesa ETL je priprava podrobnega načrta procesa ETL. Plan predstavimo v obliki grafičnega diagrama, ki zajema podrobni podatkovni tok in potrebne transformacije. Omeniti velja, da moderna orodja podatkovne integracije omogočajo snovanje procesa ETL z grafičnimi elementi, kar poveže načrtovanje in realizacijo v enotno opravilo. Prednosti orodij podatkovne integracije proti ročnemu kodiranju smo navedli v začetku 5. poglavja.

7.5. Aplikativna sled

Pri aktivnosti načrtovanja uporabniškega okolja s pomočjo orodij poslovnega obveščanja, ki smo jih obravnavali v 6. poglavju, se bomo osredotočili na izdelavo poročil, ki so lahko vključena v že omenjene nadzorne plošče. Pri načrtovanju je ključnega pomena, da ne zanemarimo mnenj končnih uporabnikov sistema. Le na ta način lahko namreč zagotovimo zadovoljitev njihovih dejanskih informacijskih potreb. Preko njihovega sodelovanja pri pripravi sistema dosežemo tudi, da bodo imeli uporabniki občutek soustvarjanja sistema, kar je dodatno zagotovilo, da bodo kasneje sistem tudi dejansko uporabljali.

Načrtovanje poročil izvedemo preko naslednjih opravil:

- določitev kandidatov za poročila
- določitev prioritet za posamezna poročila
- določitev načina dostopa do poročil za posamezne skupine uporabnikov
- določitev standardov pri izdelavi poročil
- priprava testnih poročil
- pregled testnih poročil s strani končnih uporabnikov

Pri implementaciji poročil je potrebno biti posebno pozoren na testiranje in preverbo pravilnosti rezultatov. Napačni rezultati lahko namreč pri končnih uporabnikih povzročijo dvom o kakovosti celotnega sistema poslovnega obveščanja, kar lahko ogrozi uspešnost celotnega projekta.

7.6. Namestitev sistema

Pred namestitvijo sistema je potrebno opraviti opravila, ki zajemajo:

- preverbo ustreznosti računalnikov končnih uporabnikov
- preverbo pravilnosti rezultatov
- namestitev aplikacij poslovnega obveščanja na računalnikih končnih uporabnikov
- izobraževanje končnih uporabnikov
- priprava strategije podpore končnim uporabnikom

Preverbe ustreznosti strojne, mrežne in systemske programske opreme na računalnikih končnih uporabnikov ne velja podcenjevati. Pri velikih sistemih lahko namreč ugotovitev, da računalniki ne zadoščajo potrebam sistema poslovnega obveščanja, podaljša potreben čas za namestitev sistema.

Pravilnost podatkov je potrebno preverjati v vseh stopnjah transformacij procesa ETL. Kljub temu se določene napake lahko pojavijo šele ob izobraževanju ali nudenju podpore končnim uporabnikom. Pri ugotovitvi neskladij med sistemom poslovnega obveščanja in izvirnim podatkovnim virom naletimo na tri možnosti:

- napačna je informacija v sistemu za poslovno obveščanje
- informacija v sistemu za poslovno obveščanje ima drugačen pomen, kot sorodna informacija v izvirnem sistemu
- napačna je informacija v izvirnem sistemu

Pomembno je, da se v čim večji meri izognemo prvi možnosti, saj lahko to, kot že rečeno, pri uporabnikih omaja kredibilnost sistema.

Izobraževanje končnih uporabnikov zajema tako seznanitev z uporabniškim vmesnikom, kot tudi pravilno razumevanje in interpretacijo podatkov. Zaradi različnega znanja uporabnikov, področij dela in zahtevnosti uporabniškega okolja naj bo izobraževanje prilagojeno posameznim skupinam uporabnikov in izvajano ločeno. Pomembno je, da se izobraževanje izvede šele, ko v sistemu ne zaznavamo večjih pomanjkljivosti.

Strategija podpore uporabnikom je v veliki meri odvisna od njihovega predhodnega znanja in zahtevnosti uporabniškega okolja. Med IT strokovnjaki in poslovnimi uporabniki naj se jasno razmejijo naloge. Na ta način se izognemo v praksi pogostega prelaganja nalog na IT oddelek. Stopnja samostojnosti poslovnih uporabnikov naj bo torej dogovorjena, saj lahko le tako zagotovimo učinkovito podporo uporabnikom, ki bo zadovoljila pričakovanja tako poslovnih uporabnikov, kot tudi vodstva združbe.

Ko uspešno zaključimo z navedenimi opravili, se lahko lotimo dejanske namestitve sistema. Pri velikih sistemih izvedemo še t.i. beta oziroma pilotno namestitev, vanjo pa vključimo le izbrane uporabnike. Na ta način si pred končno namestitvijo zagotovimo še dodatno kontrolo pravilnosti delovanja sistema.

7.7. Vzdrževanje in nadgradnje sistema

V primerjavi s klasičnimi projekti razvoja programske opreme je pomembnost vzdrževanja in nadgradenj sistema poslovnega obveščanja nekoliko večja. To izvira iz dejstva, da gre, kot že nekajkrat omenjeno, pri sistemih za poslovno obveščanje za nikoli dokončani projekt. Aktivnost vzdrževanja sistema poslovnega obveščanja delimo na naslednja opravila:

- zagotavljanje podpore poslovnim uporabnikom
- zagotavljanje izobraževanja poslovnih uporabnikov
- zagotavljanje učinkovitega delovanja sistema
- spremljava uporabe sistema
- promoviranje sistema

Zagotavljanje podpore poslovnim uporabnikom takoj po namestitvi sistema je pomembna, ker bi nezadovoljstvo zaradi napak ali težav zaradi nezadostnega znanja pri uporabi sistema lahko privedlo do odpora pri njegovi uporabi. Zadovoljstvo uporabnikov lahko zagotovimo s hitrim odzivom na težave in osebnim pristopom do uporabnikov. V velikih sistemih se priporoča sestava posebne skupine za podporo uporabnikom ali pa določitev terminov, v katerih so IT strokovnjaki na voljo za reševanje težav poslovnih uporabnikov.

Poleg podpore uporabnikom je po namestitvi sistema pomembno tudi njihovo nadaljnje izobraževanje. Izobraževanja je moč izvajati v različnih oblikah:

- napredno izobraževanje
- ponovitev osnovnih izobraževanj
- izobraževanja za novo zaposlene
- redni seminarji
- izobraževanja preko spletnega portala

Učinkovito delovanje sistema zagotavljamo preko treh vrst nalog. Najosnovnejša med njimi je nadzorovanje učinkovitega delovanja programske, strojne in omrežne opreme. Gre za naloge, kot so na primer ugotavljanje ozkih grl v omrežju, defragmentacija diskov in podobno. Naslednja skupina nalog zajema optimizacijo sistema, kjer je poudarek na že omenjenih konceptih agregatov, indeksov in paralelizma. Dodatna optimizacija v aktivnosti vzdrževanja sistema je mogoča zaradi povratnih informacij o dejanski uporabi sistema. Tretja in hkrati zadnja skupina nalog pa zajema vzpostavitev organizacijskih pravil, ki zagotavljajo učinkovito zaznavanje sprememb v izvornih podatkih sistema. Če se na spremembe strukture podatkov ne odzovemo s primernimi spremembami sistema namreč tvegamo, da se proces ETL ne bo uspešno izvedel ali pa, da bodo končni podatki sistema napačni.

V primeru, da so bili za uspešnost vzpostavitve sistema za poslovno obveščanje določeni merljivi kriteriji, je približno mesec dni po namestitvi sistema pravi čas za ugotovitev njihove izpolnitve. Poleg merljivih kriterijev se priporoča tudi ugotavljanje splošnega zadovoljstva pri uporabi sistema s strani poslovnih uporabnikov. Naslednji vidik spremljave uporabe sistema je zaznava poslovnih odločitev, ki delno ali v celoti temeljijo na informacijah sistema za poslovno obveščanje. Na podlagi tako zbranih odločitev je mogoče okvirno ovrednotiti vrednost sistema in izračunati čas povrnitve investicije.

Sistemi poslovnega obveščanja se od ERP sistemov bistveno razlikujejo v več pogledih. Eden od njih je tudi resnost posledic njihove neuporabe. V primeru neuporabe ERP sistema namreč zastanejo poslovni procesi v združbi, česar si uporabniki nikakor ne morejo privoščiti. Neuporaba sistema poslovnega obveščanja pa po drugi strani nima tako drastičnih neposrednih posledic. Iz tega razloga je promocija sistema in spodbujanje njegove uporabe tudi dolgo po namestitvi sistema izrednega pomena.

Zaradi narave sistemov za poslovno obveščanje je potrebno pobude za nadgradnjo sistema razumeti kot potrditev uspešnosti projekta, saj te pričajo o tem, da ga poslovni uporabniki uspešno uporabljajo. Glavna vzroka za večje nadgradnje sistema sta razširitev sistema na nova poslovna področja in zajem novih izvornih podatkov v združbi.

Z namenom spremljanja zahtev po večjih nadgradnjah sistema poslovnega obveščanja se priporoča vzpostavitev posebne skupine, ki jo sestavljajo tako IT strokovnjaki, kot tudi poslovni uporabniki. Glava naloga skupine je določitev prioritet za večje nadgradnje sistema, pri čemer je potrebno tehtati med učinkom, ki ga bo razširitev sistema imela na poslovne odločitve, ter izvedljivostjo projekta v razumnih časovnih okvirih.

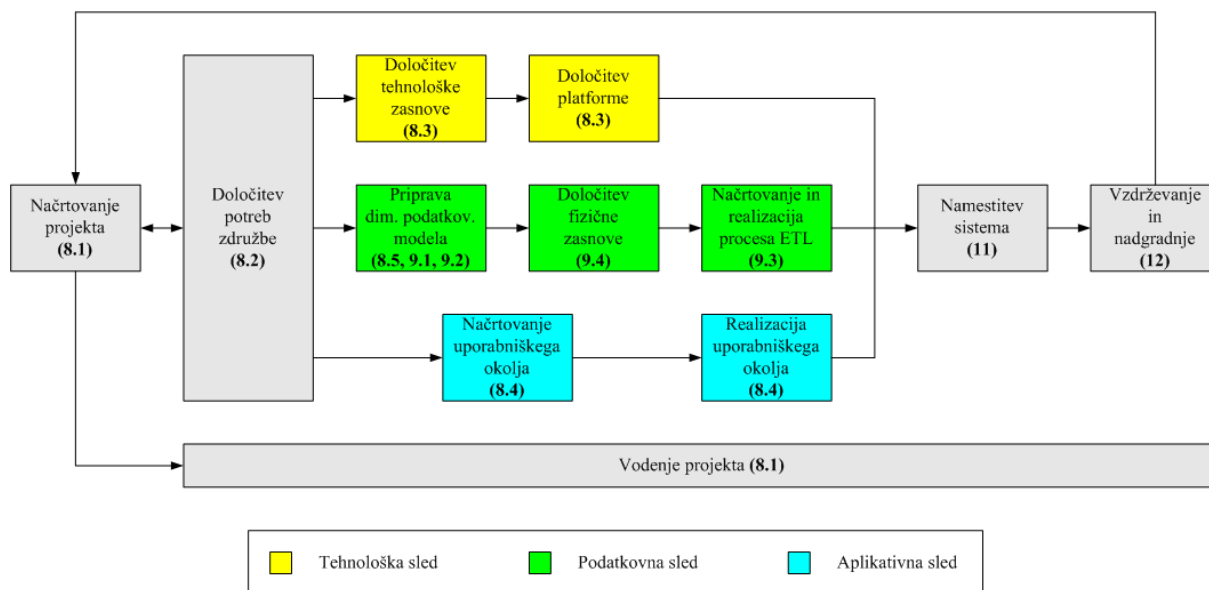
Želje po manjših nadgradnjah sistema, pod katere štejemo spremembe v okviru obstoječih področnih podatkovnih skladišč, uporabniki sporočajo na bolj neformalne načine, na primer ustno ali preko elektronske pošte. Kljub temu, pa želj po manjših nadgradnjah sistema ne velja preslišati, saj lahko večje število manjših popravkov pomembno izboljša učinkovitost sistema.

Za konec navedimo še zadnje priporočilo. Enako kot pri realizaciji osnovnega sistema poslovnega obveščanja velja tudi pri večjih nadgradnjah slediti metodologiji *Življenjskega cikla*.

8. Podatkovno skladišče CPK d.d.

V nadaljevanju naloge bomo predstavili realizacijo podatkovnega skladišča CPK d.d. na osnovah metodologije *Življenjskega cikla* [13]. Določene aktivnosti bomo obravnavali v tekočem poglavju, saj sovpadajo s celotnim podatkovnim skladiščem, nekatere pa bomo predstavili na primeru področnega podatkovnega skladišča *Glavna knjiga* v poglavju, ki sledi.

Na sliki 13 so predstavljene aktivnosti metodologije *Življenjskega cikla* z označenimi poglavji in podpoglavji, v katerih bomo na primeru podatkovnega skladišča CPK d.d. posamezne aktivnosti podrobneje predstavili.



Slika 13. Pregled obravnavanja posameznih aktivnosti po metodologiji *Življenjskega cikla*

8.1. Načrtovanje in vodenje projekta

Prvo opravilo, ki ga je bilo potrebno izvesti še pred začetkom načrtovanja projekta, je ugotavljanje zrelosti združbe za realizacijo sistema poslovnega obveščanja. Ugotovljeno je bilo, da združba izpolnjuje vseh pet kriterijev predstavljenih v podpoglavju 7.1.:

- projekt uživa podporo vodstva združbe, ki je za projekt zagotovilo finančna sredstva
- zagotavljanje kvalitetnejših informacij za poslovno odločanje sovпада s strategijo združbe
- oddelek za informatiko in poslovni uporabniki tesno sodelujejo
- poslovni uporabniki v nekaterih oddelkih si želijo enostavnejši dostop do kakovostnejših informacij
- izvorni podatki ERP sistema *Navision* zagotavljajo uspešno realizacijo sistema

Obseg projekta je bil določen z izborom področnih podatkovnih skladišč in vrstnim redom njihove realizacije:

1. Glavna knjiga
2. Saldakonti

3. Materialno poslovanje
4. Evidenca dela
5. Projektno vodenje
6. Plače
7. Dokumentni sistem

Kot je razvidno iz določitve obsega projekta, realizacija sistema poslovnega obveščanja v veliki meri poteka po pristopu od spodaj navzgor. Za omenjeni pristop se kot pomanjkljivost najpogosteje izpostavlja težavno zagotavljanje konsistentnosti podatkov, česar pa zaradi enotnega in kakovostnega podatkovnega vira, ki ga predstavlja ERP sistem *Navision* v našem projektu ni zaznati. S pristopom od spodaj navzgor sledimo želji po hitrih rezultatih z relativno nizkimi stroški.

V okviru izdelave načrta projekta je bil izbran naziv projekta ter določen okviren terminski plan. Ker je bilo za posamezno področno podatkovno skladišče že v začetku predvideno veliko število iteracij, je okviren terminski plan določen le za prvo iteracijo posameznega izmed njih.

Izziv vodenja projekta predstavlja predvsem aktivno vključevanje poslovnih uporabnikov v testiranja, od katerih se poleg preverjanja pravilnosti rezultatov pričakuje tudi podajanje pripomb in želj po novih funkcionalnostih. Te se v primeru ugotovljene upravičenosti realizirajo ob naslednji iteraciji. Vodja projekta je v okviru aktivnosti vodenja projekta zavezan k poročanju vodstvu združbe o poteku projekta. Na rednih sestankih se vodstvo združbe seznaja z novimi funkcionalnostmi in konkretnimi rezultati.

8.2. Določitev potreb združbe

Potrebe združbe ugotavljamo preko poslovnih uporabnikov. Za njihovo določitev poznamo dve tehniki in sicer na podlagi intervjujev ali vodenih sestankov. Za potrebe našega projekta smo se odločili za določitev potreb na podlagi intervjujev, za kar smo se odločili iz sledečih razlogov:

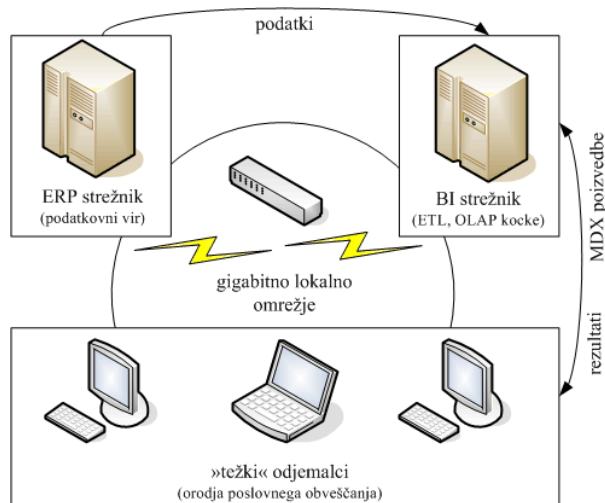
- za določitev potreb so bili izbrani predvsem vodstveni kadri, ki jih z intervjuji časovno manj obremenimo
- v začetnih iteracijah razvoja bo sistem uporabljalo omejeno število poslovnih uporabnikov
- z individualnim pristopom si zagotovimo več povratnih informacij

Zbrane potrebe združbe je bilo potrebno preučiti s stališča možnosti njihove zagotovitve, upoštevajoč razpoložene podatkovne vire. Ugotovljeno je bilo, da je potrebam možno v veliki meri ugoditi že na podlagi obstoječih podatkov, ki se zbirajo v ERP sistemu *Navision*. V kasnejših iteracijah so za zagotovitev določenih potreb vseeno predvidene manjše dopolnitve ERP sistema.

Kot primer dopolnitve ERP sistema zaradi specifične potrebe poslovnih uporabnikov navedimo združevanje poslovnih partnerjev na podlagi pravne oblike. Ker v osnovi tega podatka v ERP sistemu nismo vodili, je bilo potrebno evidenco vzpostaviti naknadno.

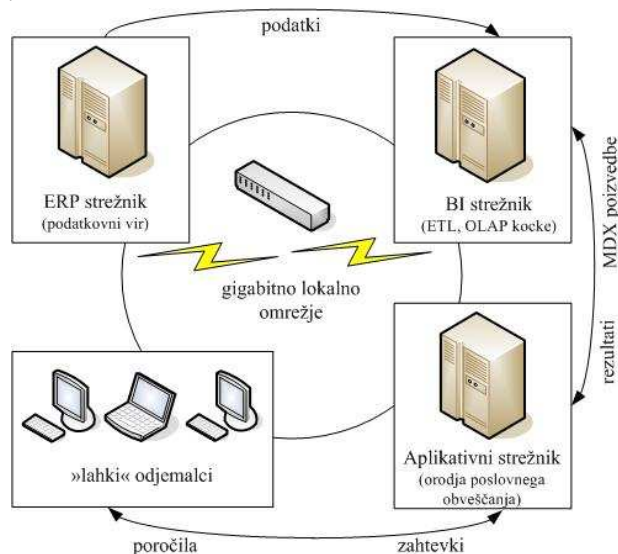
8.3. Določitev tehnološke zasnove in platforme

V okviru tehnološke zasnove je bil pripravljen okvirni načrt tehnološke arhitekture in infrastrukture. V začetni fazi je bila predvidena uporaba dvonivojske arhitekture, ki je predstavljena na sliki 14.



Slika 14. Dvonivojska arhitektura sistema poslovnega obveščanja CPK d.d.

Po načrtovani uvedbi specializiranega orodja poslovnega obveščanja pa predvidevamo prehod na tronivojsko arhitekturo, kjer bo naloge težkih odjemalcev prevzel aplikativni strežnik, kar je prikazano na sliki 15.

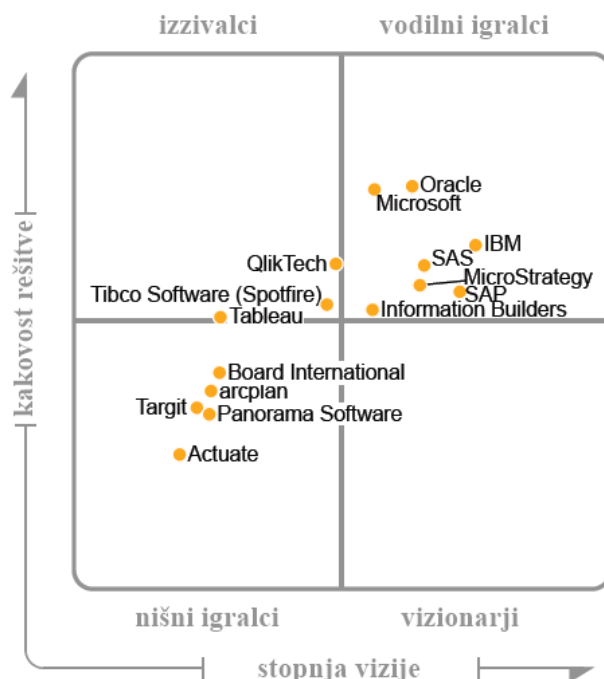


Slika 15. Predvidena tronivojska arhitektura sistema poslovnega obveščanja CPK d.d.

Izbor platforme ni bil izveden na sistematičen način. Za platformo združbe *Microsoft* smo se odločili predvsem zaradi obstoječih znanj in izkušenj, kar je bilo odločilno predvsem zaradi odločitve, da bomo skušali projekt v veliki meri realizirati z lastnim znanjem. Poleg tega izbrana platforma pokriva vse funkcionalnosti, potrebne za realizacijo sistema poslovnega obveščanja. V primeru izbora različnih ponudnikov za podporo posameznim komponentam sistema namreč tvegamo, da bomo naleteli na težave zaradi pomanjkljive medsebojne kompatibilnosti.

Na sliki 16 je prikazan *Magični kvadrant platform poslovnega obveščanja* po raziskavi združbe *Gartner Group* iz leta 2010 [7], kjer so pomembnejši igralci na tržišču razvrščeni med:

- nišne igralce, ki so specializirani le za določeno področje sistemov poslovnega obveščanja
- vizionarje, za katere so značilne odpre in fleksibilne arhitekture, a na zagotavljajo popolne funkcionalnosti na vseh področjih
- izzivalce, katerih rešitve so primerne le za specifične potrebe
- vodilni igralci, ki so na globalnem nivoju najmočnejši na tržišču



Slika 16. Magični kvadrant platform poslovnega obveščanja združbe Gartner group

Po omenjeni raziskavi se prednosti *Microsofta* kažejo v:

- učinkoviti povezavi z ostalimi produkti, kot so *Microsoft Office* in *Microsoft SharePoint*
- konkurenčni ceni glede na ponujene funkcionalnosti
- trendu večanja tržnega deleža
- dvakrat pogostejši uporabi OLAP funkcionalnosti glede na konkurenčne rešitve

Najpomembnejše pomanjkljivosti glede na konkurenco so:

- nizka stopnja povezanosti med sistemom za poslovno obveščanje in *Microsoftovimi* ERP sistemi
- odsotnost centralnega repozitorija metapodatkov na nivoju celotnega sistema poslovnega obveščanja
- počasno razvijanje novih funkcionalnosti
- počasno odzivanje na novosti na trgu

8.4. Načrtovanje in realizacija uporabniškega okolja

Ker se v začetni fazi projekta kot orodje poslovnega obveščanja uporablja *Microsoft Excel*, je bilo načrtovanje uporabniškega okolja omejeno na določitev dinamičnih poročil. Seznam poročil, katerega izsek za področje *Glavna knjiga* prikazujemo v preglednici 4, vsebuje naslednje podatke:

- kocka OLAP, na katerega se poročilo povezuje
- naziv poročila
- opis vsebine poročila
- prioriteta za realizacijo poročila na intervalu od 1 do 5

Naziv kocke OLAP	Naziv poročila	Opis vsebine	Prioriteta
Glavna knjiga	Stroški po kontih in PC	Stroški po kontih in prihodkovnih centrih	3
Glavna knjiga	Prihodki po PC	Deleži prihodkov po prihodkovnih centrih	3
Glavna knjiga	KPI	Temeljni kazalniki poslovanja	4
Glavna knjiga	Prilivi in odlivi	Prilivi in odlivi združbe po vrstah	5
Glavna knjiga	Uspešnost SM	Uspešnost stroškovnih mest	3
Glavna knjiga	Uspešnost SM po vodjih	Uspešnost stroškovnih mest po vodjih gradbišč	3

Preglednica 4. Izsek iz seznama poročil

Poudariti velja, da se glede na potrebe poslovnih uporabnikov seznam ter posledično repozitorij poročil sproti dopolnjuje. Dodatno se število realiziranih poročil povečuje tudi zaradi poslovnih uporabnikov, ki poročila samostojno ustvarjajo in jih ponudijo na voljo tudi ostalim uporabnikom.

Pri realizaciji poročil je posebna skrb namenjena pravilnosti rezultatov, ki se preverja tako s strani razvijalcev kot tudi poslovnih uporabnikov.

8.5. Priprava matrike podatkovnega skladišča

Na preglednici 5 je prikazana matrika podatkovnega skladišča za trenutno realizirana področja.

Dimenzija / Področno podatkovno skladišče	Datum	Stik	Artikel	PC	SM	SN	Konto GK	Prilivi odlivi	Vrsta postavke	Vrsta post. artikla	Oznaka terjatve	Oznaka obveznosti	Skladišče	Spl. knj. sk. proizvod.	Delavec	Obračun	Vrsta plačila
Glavna knjiga	X	X		X	X	X	X	X	X					X			
Saldakonti	X	X		X	X	X					X	X					
Materialno poslovanje	X	X	X	X	X	X				X			X				
Evidenca dela	X			X	X	X									X	X	X

Preglednica 5. Matrika podatkovnega skladišča

9. Področno podatkovno skladišče Glavna knjiga

Pri realizaciji področnih podatkovnih skladišč smo, kot že omenjeno, uporabili pristop v 7. poglavju predstavljene metodologije *Življenjskega cikla*, ki smo jo zaradi manjšega obsega našega projekta nekoliko prilagodili. Kot določeno v podpoglavju 8.3. smo za realizacijo sistema uporabili *Microsoft SQL Server 2008*, ki zagotavlja avtomatsko beleženje metapodatkov in v veliki meri združuje aktivnosti načrtovanja in realizacije. Iz tega razloga smo lahko metodološko predpisane dokumente na določenih mestih zamenjali s preglednicami in diagrami iz razvojnega okolja.

Pregled realizacije področnih podatkovnih skladišč bomo začeli s področnim podatkovnim skladiščem *Glavna knjiga*, ki poslovnim uporabnikom omogoča izvajanje podrobnih analiz:

- bilance stanja
- izkaza uspeha
- prilivov in odlivov
- temeljnih kazalnikov poslovanja združbe

Poslovna terminologija v področnem podatkovnem skladišču *Glavna knjiga*, kot tudi ostalih področnih podatkovnih skladiščih v veliki meri sledi terminologiji ERP sistema *Navision*. S takim pristopom želimo zmanjšati verjetnost napačnega razumevanja rezultatov, saj je terminologija poslovnim uporabnikom dobro poznana.

9.1. Izvorni sistem

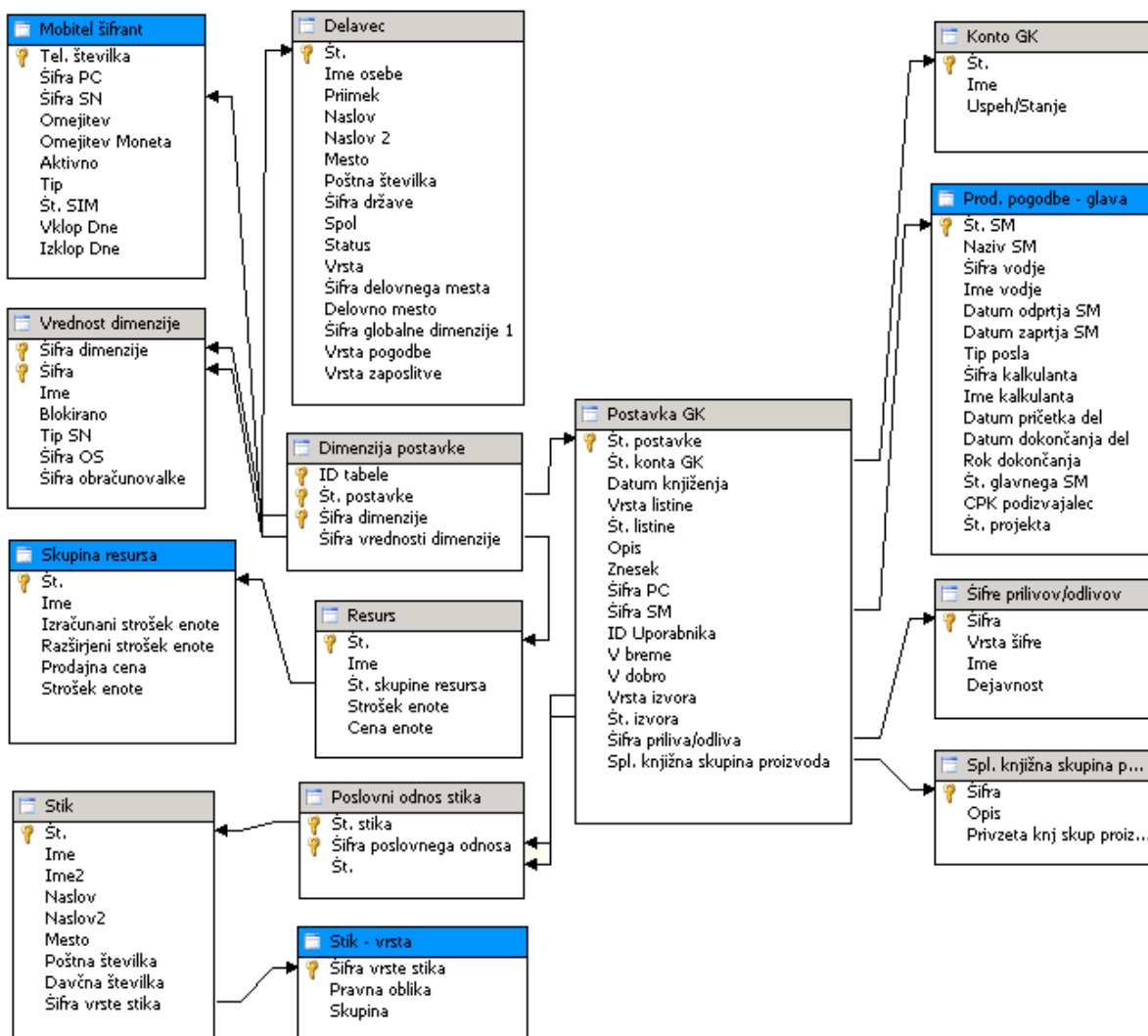
Izvorni sistem področnega podatkovnega skladišča *Glavna knjiga* v celoti predstavlja ERP sistem *Navision*. Iz tega razloga je kakovost podatkov iz sintaktičnega vidika v celoti zagotovljena, saj so napake te vrste onemogočene že z omejitvami na nivoju podatkovne baze izvirnega sistema. Zaradi načina evidentiranja knjigovodskih podatkov, ki predstavlja osnovno funkcionalnost ERP sistema, je zagotovljen tudi visok nivo kakovosti podatkov iz vidikov nedvoumnosti, konsistentnosti in popolnosti. Pričakujemo, da se bo z uporabo sistema poslovnega obveščanja za področje glavne knjige povečala kakovost podatkov iz vidika pravilnosti, torej skladnosti z realnostjo, saj se bo zaradi učinkovite analize podatkov skrajšal čas, ki je potreben za odkrivanje odstopanj podatkov od realnosti.

Za potrebe področnega podatkovnega skladišča *Glavna knjiga* smo večino podatkov črpali iz področij, ki jih pokriva že osnovna funkcionalnost ERP sistema *Navision*, nekatere podatke pa smo pridobili iz modulov, ki smo jih razvili namensko za potrebe združbe.

Analiziranje knjigovodskih podatkov je za združbo izrednega pomena. To lahko utemeljimo z dejstvom, da ERP sistem *Navision*, ki ga uporabljajo v številnih srednje velikih združbah v Sloveniji in po svetu, že v osnovni funkcionalnosti omogoča pregled podatkov izključno področja glavne knjige na načine, ki spominjajo na sisteme poslovnega obveščanja. Tako iz vidika hitrosti delovanja, kot tudi različnih možnosti uporabe, pa so ti pregledi v primerjavi z modernimi sistemi poslovnega obveščanja zelo omejeni. Tudi iz tega razloga se nam odločitev, da kot prvo področno podatkovno skladišče podrobno predstavimo prav to področje, zdi utemeljena.

Na sliki 17 je prikazan entitetno relacijski diagram izvornega sistema, pri čemer so standardne tabele ERP sistema *Navision* označene s sivo barvo, tabele, ki pokrivajo funkcionalnosti, razvite za potrebe združbe, pa z modro barvo.

Posebnost sistema *Navision* je poimenovanje tabel in atributov tako v angleškem, kot tudi lokalnem (slovenskem) jeziku. Pri prikazu našega ER diagrama smo seveda uporabili slovenščino, velja pa poudariti, da se je v SQL poizvedbah potrebno sklicevati na originalne (angleške) nazive.



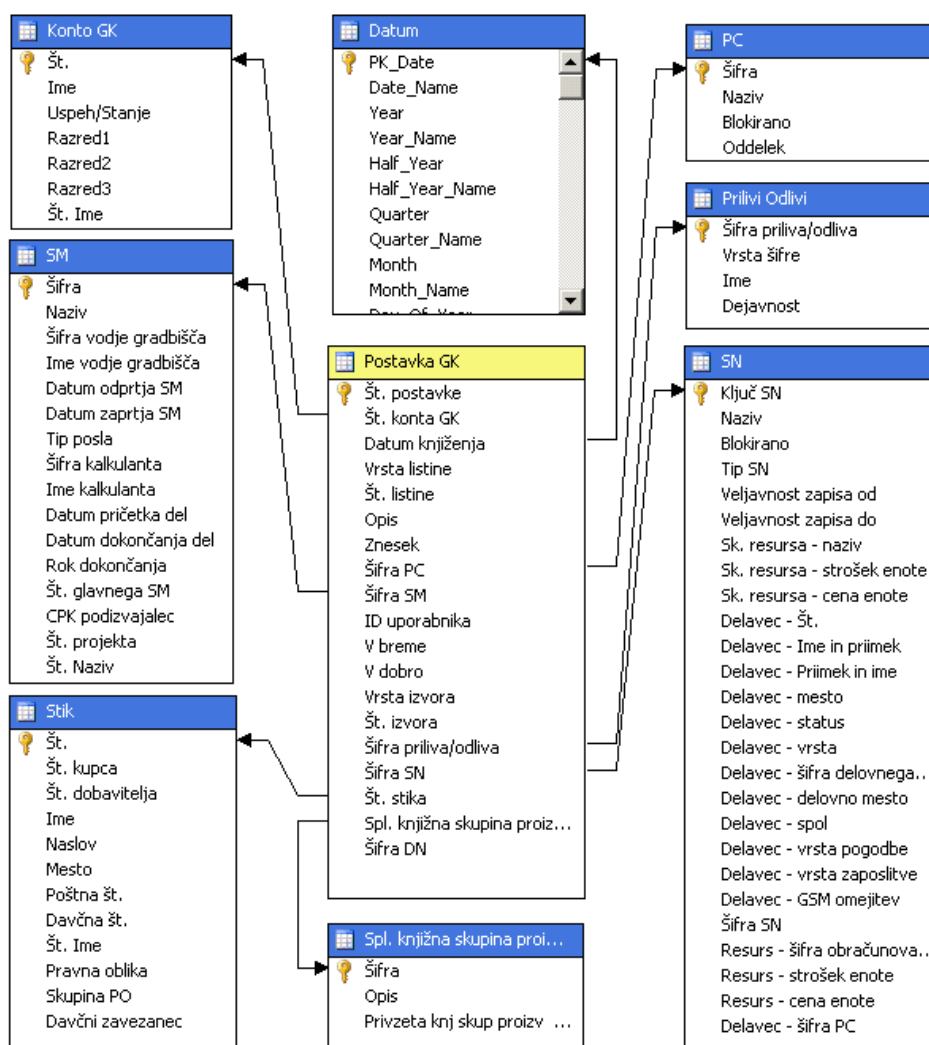
Slika 17. ER diagram izvornega sistema za področno podatkovno skladišče Glavna knjiga

Pri izboru notacije ER diagrama smo sledili notaciji, uporabljeni v *Microsoft SQL Server Analysis Services 2008* (SSAS). Puščice, ki na diagramu predstavljajo povezavo med posameznimi tabelami predstavljajo kardinalnost N:1 v smeri puščice.

9.2. Priprava dimenzijskega podatkovnega modela

Pred predstavitvijo dimenzijskega podatkovnega modela področnega podatkovnega skladišča *Glavna knjiga*, omenimo še preglednico *Pregled toka podatkov*. V preglednici za vsak atribut posamezne tabele dimenzijskega podatkovnega modela zabeležimo njegov izvor, potrebno transformacijo v opisni obliki ter ponor v dimenzijskem podatkovnem modelu. Z namenom učinkovitejše predstavitve bomo *Pregled toka podatkov* predstavili skupaj z grafičnim prikazom transformacij v podpoglavju 9.3.

Tudi dimenzijski podatkovni model področnega podatkovnega skladišča *Glavna knjiga* je predstavljen z notacijo, uporabljeno v SSAS (slika 15). Dodatno je v preglednici 6 za vsak atribut tabele dejstev *Postavka GK* naveden še tip atributa, kot to narekuje *Podrobni diagram tabele dejstev* metodologije *Življenjskega cikla*.



Slika 18. Dimenzijski podatkovni model področnega podatkovnega skladišča Glavna knjiga

Kot je razvidno iz slike 18, smo pri dimenzijskem podatkovnem modelu *Glavna knjiga* uporabili podatkovno strukturo vrste zvezdna shema. Zrno tabele dejstev *Postavka GK* predstavlja posamezni zapis glavne knjige.

Naziv atributa	Tip atributa
Št. postavke	Cljuč tabele dejstev
Št. konta GK	Cljuč dimenzijske tabele
Datum knjiženja	Cljuč dimenzijske tabele
Vrsta listine	Degenerirana dimenzija
Št. listine	Ni v uporabi
Opis	Ni v uporabi
Znesek	Osnovno dejstvo
Šifra PC	Cljuč dimenzijske tabele
Šifra SM	Cljuč dimenzijske tabele
ID uporabnika	Degenerirana dimenzija
V breme	Osnovno dejstvo
V dobro	Osnovno dejstvo
Vrsta izvora	Degenerirana dimenzija
Št. izvora	Ni v uporabi
Šifra priliva / odliva	Cljuč dimenzijske tabele
Šifra SN	Cljuč dimenzijske tabele
Št. stika	Cljuč dimenzijske tabele
Spl. knjižna skupina proizv.	Cljuč dimenzijske tabele
Šifra DN	Ni v uporabi

Preglednica 6. Podrobni diagram tabele dejstev Glavna knjiga

Za lažje razumevanje vsebine področnega podatkovnega skladišča *Glavna knjiga* si na sliki 19 pogledjmo nekaj konkretnih zapisov tabele dejstev *Postavka GK*.

Št. postavke	Št. konta GK	Datum knjiženja	Vrsta listine	Znesek	Šifra PC	Šifra SM	ID uporabnika	V breme	V dobro	Vrsta izvora	Št. izvora	Šifra priliva...	Šifra SN	Št. stika	Spl. knjižna ...
2772929	320010	2009-02-04		-662,1	34	-	KOVACEVICM	0	662,1			NULL	3149	-	-
2772930	404020	2009-02-04		662,1	34	-	KOVACEVICM	662,1	0			NULL	3149	-	-
2772931	321010	2009-02-04		662,1	34	-	KOVACEVICM	662,1	0			NULL	3149	-	-
2772932	329010	2009-02-04		-662,1	34	-	KOVACEVICM	0	662,1			NULL	3149	-	-
2773118	310010	2009-02-04		-11,96	23	-	OTONICARA	0	11,96			NULL	3018	-	-
2773119	403010	2009-02-04		11,96	23	-	OTONICARA	11,96	0			NULL	3018	-	-
2773700	412010	2009-01-22	Račun	34,98	33	-	NIKOLICM	34,98	0	Dobavitelj	212	NULL	3148	22007	STORITV-SS
2773701	160010	2009-01-22	Račun	7	33	-	NIKOLICM	7	0	Dobavitelj	212	NULL	3148	22007	-
2773702	220010	2009-01-22	Račun	-41,98	33	-	NIKOLICM	0	41,98	Dobavitelj	212	NULL	3148	22007	-
2773728	412010	2009-01-26	Račun	34,98	33	-	NIKOLICM	34,98	0	Dobavitelj	212	NULL	2760	22007	STORITV-SS
2773729	160010	2009-01-26	Račun	7	33	-	NIKOLICM	7	0	Dobavitelj	212	NULL	2760	22007	-

Slika 19. Primeri zapisov tabele dejstev Postavka GK

9.2.1. Pregled dimenzijskih tabel

Tako kot dimenzijski podatkovni model bomo tudi podrobni pregled dimenzijskih tabel predstavili s pomočjo posnetkov zaslona iz aplikacije SSAS. Na njih so razvidne hierarhije, ki smo jih določili glede na hierarhično povezanost atributov znotraj dimenzijske tabele in predvidene potrebe poslovnih uporabnikov. Za vsako dimenzijsko tabelo bomo poleg omenjenih posnetkov zaslona prikazali še preglednico s sledečimi lastnostmi atributov:

- naziv atributa
- opis atributa
- primer vrednosti atributa
- število različnih vrednosti (izračunano na določen datum)

- tip počasi se spreminjajoče dimenzije

S prikazom hierarhij in dodatne preglednice bomo predstavili vse podatke, ki jih preko *Podrobnega diagrama dimenzijskih tabel* predvideva metodologija *Življenjskega cikla*. Za potrebe naloge, bomo predstavili štiri od skupaj osmih dimenzijskih tabel.

Dimenzijska tabela *Datum*, ki bo prva deležna podrobne obravnave, ima v podatkovnem skladišču posebno vlogo. Poleg tega, da se nahaja v vsakem področnem podatkovnem skladišču, praviloma nima podatkovnega vira, kar pomeni, da je potrebno njeno vsebino vpisati ročno ali napolniti programsko. Funkcionalnost polnjenja datumskih dimenzijskih tabel je v SSAS za angleški jezik podprto že v osnovi, kar bomo za dimenzijsko tabelo *Datum* s pridom izkoristili, ter se tako izognili dodatnemu programiranju.

V našem primeru smo dimenzijsko tabelo datum napolnili z zapisi, ki zajemajo obdobje od leta 2003 do leta 2018. Med mnogimi atributi dimenzijske tabele *Datum*, ki jih podpira SSAS smo uporabili le najosnovnejše. V primeru, da bi poslovni uporabniki izkazali potrebo po dodatnih atributih ali hierarhijah, bi bilo potrebno strukturo in vsebino dimenzijske tabele *Datum* dopolniti, kar pa v SSAS ne predstavlja težave. Primer takšne razširitve bi lahko bil na primer dodatni atribut *Dan v tednu*.

Dimenzijska tabela *Datum* se v področnem podatkovnem skladišču *Glavna knjiga*, kot tudi vseh ostalih področnih podatkovnih skladišč uporablja za omejitev rezultatov na izbrana časovna obdobja.



Slika 20. Hierarhija dimenzijske tabele Datum

Naziv atributa	Opis atributa	Primer vrednosti atributa	Št. različnih vrednosti	Tip PSSD
Year	Leto	calendar 2010	16	/
Half year	Polletje	semester 1, 2010	32	/
Quarter	Četrletje	quarter 1, 2010	64	/
Month	Mesec	januar 2010	192	/
Date	Datum	12.1.2010	5844	/

Preglednica 7. Pregled atributov dimenzijske tabele Datum

Naslednja dimenzijska tabela, ki jo bomo podrobneje predstavili, je *Konto GK*. Poslovnim uporabnikom omogoča, da omejijo rezultate glede na izbrane skupine kontov na več nivojih, kar v računovodskem žargonu imenujejo sintetika, oziroma na posamezne konte, za kar se uporablja termin analitika.

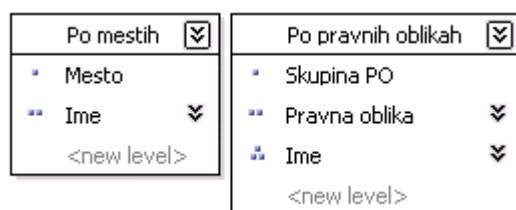


Slika 21. Hierarhija dimenzijske tabele Konto GK

Naziv atributa	Opis atributa	Primer vrednosti atributa	Št. različnih vrednosti	Tip PSSD
Št	Številka konta	411010	1641	/
Ime	Ime konta	Stroški gradbenih storitev	1641	1
Uspeh / stanje	Oznaka, ali gre za konto bilance stanja ali izkaza uspeha	Izkaz uspeha	2	1
Razred 1	Enoštevilčna skupina konta	4	10	1
Razred 2	Dvoštevilčna skupina konta	41	72	1
Razred 3	Troštevilčna skupina konta	411	358	1
Št. Ime	Številka in ime konta	411010 Stroški gradbenih storitev	1641	1

Preglednica 8. Pregled atributov dimenzijske tabele Konto GK

Dimenzijska tabela *Stik* vsebuje seznam poslovnih partnerjev združbe. Preko nje lahko poslovni uporabniki omejijo želene rezultate za tiste postavke GK, ki podatek o poslovnem partnerju vsebujejo, kar velja za večino kontov izkaza uspeha. Podobno kot v dimenzijski tabeli *Datum*, so tudi v tabeli *Stik* v področnem podatkovnem skladišču trenutno uporabljeni le osnovni atributi. Za potrebe podatkovnega skladišča je bil v ERP sistemu dodan atribut *Vrsta stika*, glede na potrebe poslovnih uporabnikov pa bodo lahko naknadno vključeni tudi dodatni atributi iz podatkovnega vira.



Slika 22. Hierarhiji dimenzijske tabele Stik

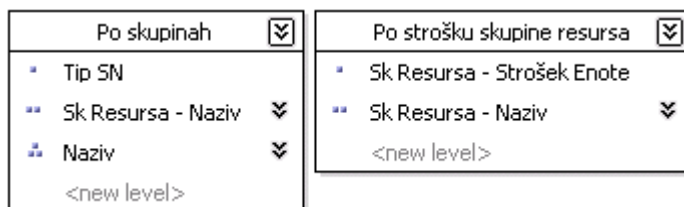
Naziv atributa	Opis atributa	Primer vrednosti atributa	Št. različnih vrednosti	Tip PSSD
Št	Številka stika	14000	5038	/
Ime	Ime stika	Lesonit d.d.	5038	1
Mesto	Mesto stika	Ilirska Bistrica	324	1
Davčni zavezanec	Oznaka, ali gre za dav. zavez.	zavezanec	3	1
Št. Ime	Številka in ime stika	14000 Lesonit d.d.	5038	1
Pravna oblika	Pravna oblika	d.d.	14	1
Skupina PO	Skupina pravne oblike	pravna oseba	5	1

Preglednica 9. Pregled atributov dimenzijske tabele Stik

Dimenzijska tabela stroškovnih nosilcev (SN) je zadnja dimenzijska tabela, ki bo deležna podrobne obravnave. V tabeli smo zaradi atributa *Delavec – GSM omejitev*, za katerega želimo hraniti zgodovinske vrednosti, uporabili koncept nadomestnega ključa. Podatkovni tip nadomestnega ključa je celo število (integer), njegova vrednost pa se določi avtomatsko, na nivoju podatkovne baze. Postopek iskanja nadomestnega ključa za ustvarjanje povezave med tabelo dejstev in dimenzijsko tabelo SN je opisan v razdelku 9.3.7.

Posamezen zapis v dimenzijski tabeli SN predstavlja zapis o stroju ali delavcu, kar je razvidno iz atributa *Tip SN*. Ker gre za zelo različna poslovna subjekta, so za vsakega izmed njih definirani tudi zanj lastni atributi. Atributi, ki za določen zapis ne pripadajo primernemu tipu SN, imajo za ta zapis prazno vrednost.

V dimenzijski tabeli SN smo za hierarhijo *Po strošku skupine resursa* uporabili napredno možnost diskretizacije, ki nam jo omogoča SSAS. Gre za avtomatsko ustvarjanje hierarhičnih skupin na glede na interval numeričnih vrednosti. Tak pristop nam za attribute, katerih vrednost je numerična, omogoča manjše število skupin na določenem hierarhičnem nivoju, kot, če bi ustvarili skupino za vsako enolično numerično vrednost.



Slika 23. Hierarhiji dimenzijske tabele SN

Naziv atributa	Opis atributa	Primer vrednosti atributa	Št. različnih vrednosti	Tip PSSD
Ključ SN	Nadomestni ključ	3608	1239	/
Veljavnost zapisa od	Datum začetka veljavnosti zapisa	1.1.2003	2	/
Veljavnost zapisa do	Datum konca veljavnosti zapisa	31.12.2008	2	/
Šifra SN	Poslovni ključ	30123	1238	/
Naziv	Naziv stroškovnega nosilca	Janez Novak	1238	1
Blokirano	Oznaka, ali je SN blokirano	NE	2	1
Tip SN	Oznaka tipa SN	Delavec	6	1
Delavec - Ime in priimek	Ime in priimek delavca	Janez Novak	431	1
Delavec - Priimek in ime	Priimek in ime delavca	Janez Novak	431	1
Delavec - status	Oznaka, ali je delavec aktiven	Aktiven	2	1
Delavec - spol	Spol delavca	Moški	2	1
Delavec - mesto	Mesto prebivališča delavca	Koper	48	1
Delavec - delovno mesto	Delovno mesto delavca	Vodja informatike	101	1
Delavec – GSM omejitev	Mesečna kvota	20	12	2
Sk resursa - naziv	Naziv skupine resursa	Strojnik asfaltne baze	89	1
Sk resursa – strošek enote	Strošek skupine resursa na uro	8,92	82	1

Preglednica 10. Pregled atributov dimenzijske tabele SN

Posebnost atributa *Delavec – GSM omejitev* je poleg hranjenja zgodovinskih vrednosti tudi ta, da želimo vrednosti atributa poslovnemu uporabniku predstaviti na način, kot bi se atribut nahajal v tabeli dejstev. S tem želimo poslovnemu uporabniku omogočiti prikaz vrednosti atributa iz dimenzijske tabele med rezultati poročila, torej na enak način, kot so prikazana merljiva dejstva iz tabele dejstev.

Realizacija obeh posebnosti je mogoča s posebno vrsto povezave med dimenzijo in skupino merljivih dejstev, ki jo imenujemo *Večkratna povezava* in je ena izmed možnih vrst povezave, ki jih omogoča SSAS. Vrste povezav med dimenzijami bomo obravnavali v razdelku 9.5.2.

9.3. Načrtovanje in realizacija procesa ETL

Po Kimballu je proces ETL sestavljen iz aktivnosti izbora, čiščenja, poenotenja in dostave. Kot že omenjeno v 5. poglavju, za realizacijo procesa ETL potrošimo v povprečju 70% časa celotnega projekta.

V nadaljevanju si bomo pogledali realizacijo procesa ETL področnega podatkovnega skladišča Glavna knjiga z uporabo *MS SQL Server Integration Services* (SSIS). Omeniti velja, da so pri snovanju procesa ETL v omenjenem orodju, ločnice med aktivnostmi procesa ETL, kot jih predvideva Kimball, povsem zabrisane. Snovanje procesa ETL poteka preko grafičnega vmesnika, z uporabo različnih grafičnih objektov oziroma gradnikov.

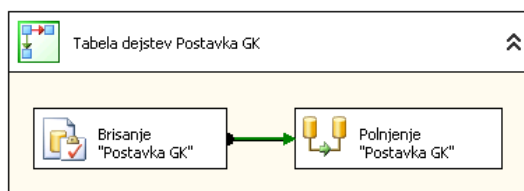
Prednosti realizacije procesa ETL z uporabo modernega, grafičnega orodja je mnogo. Poleg prednosti, naštetih v 5. poglavju, omenimo še dodatno poenostavitev, ki nam jo tak pristop ponuja. Zaradi nazorne, grafične ponazoritve procesov in avtomatskega generiranja repozitorija metapodatkov namreč ni potrebe po dodatnem, ročnem dokumentiranju procesa ETL.

9.3.1. Globalni paket

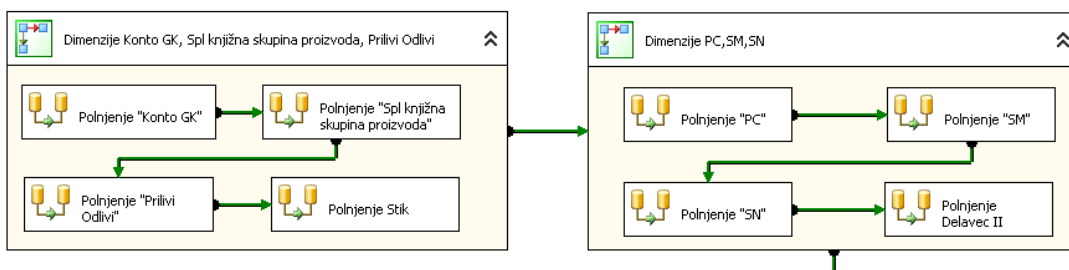
Zaradi lažjega obvladovanja kompleksnosti je snovanje procesa ETL v SSIS realizirano v več nivojih. Na najvišjem nivoju definiramo t.i. SSIS pakete, v katerih je priporočljivo določiti vsebnike. Znotraj vsebnikov (containers) določimo opravila, katerih vrsto izberemo iz repozitorija SSIS opravil. Tako za vsebnike kot tudi opravila znotraj njih je potrebno določiti zaporedje njihovega izvajanja. Za potrebe področnega podatkovnega skladišča *Glavna knjiga* smo uporabili opravila SQL in podatkovno transformacijo.

Pri realizaciji kompleksnih sistemov je po načelu »deli in vladaj« smiselno uporabiti možnost hierarhične uporabe SSIS paketov, kar z drugimi besedami pomeni, da znotraj posameznega SSIS paketa vključimo SSIS pakete nižjega nivoja.

Na sliki 24 je predstavljen SSIS paket področnega podatkovnega skladišča *Glavna knjiga*, ki je zasnovan za polnjenje tabele dejstev in vsebuje eno opravilo SQL in eno podatkovno transformacijo. Polnjenje dimenzijskih tabel za vsa področna podatkovna skladišča je zasnovano v ločenem SSIS paketu, katerega del je prikazan na sliki 25.



Slika 24. SSIS paket področnega podatkovnega skladišča Glavna knjiga



Slika 25. Izsek SSIS paketa Polnjenje dimenzij

9.3.2. Gradniki opravi

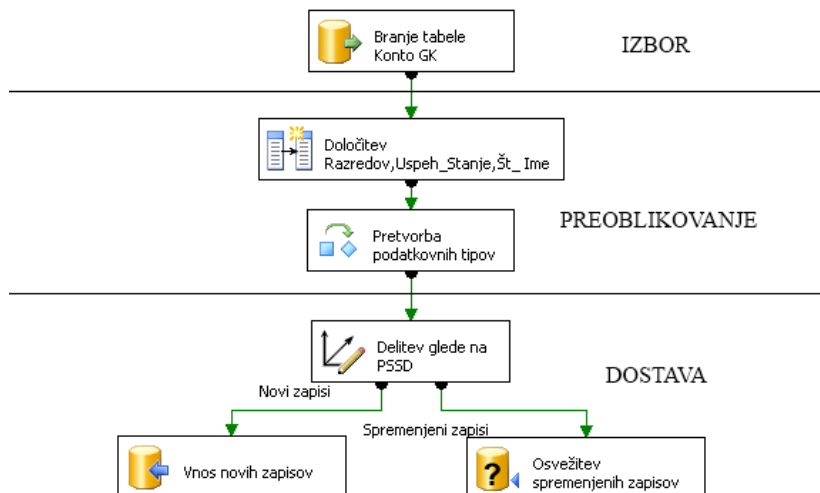
Snovanje posameznih opravi v SSIS, kot že omenjeno, poteka v grafičnem okolju s pomočjo gradnikov. Preden se poglobimo v vsebino posameznih opravi bomo v preglednici 11 izmed številnih gradnikov, ki jih vsebuje SSIS, opisali funkcionalnost tistih, ki smo jih uporabili za realizacijo opravi našega podatkovnega skladišča.

Simbol gradnika	Naziv gradnika	Opis funkcionalnosti
	Branje zapisov	Omogoča branje iz podatkovnih virov z uporabo SQL poizvedb, ki so lahko vpisane ročno ali generirane preko grafičnega vmesnika.
	Vnos zapisov	Omogoča pisanje v podatkovne baze na podlagi preslikave med atributi opravi in atributi tabele podatkovne baze.
	Vnos zapisov z uporabo parametrov	Omogoča pisanje v podatkovne baze z uporabo parametrov. Gradnik se uporablja v kombinaciji s počasi se spreminjajočimi dimenzijami.
	Preoblikovanje na nivoju zapisa	Omogoča preoblikovanje podatkov na nivoju posameznega zapisa.
	Sprememba podatkovnega tipa	Omogoča pretvorbe podatkovnih tipov atributov.
	Iskanje zapisa (lookup)	Omogoča iskanje zapisa v izbrani tabeli glede na vrednosti iskalnih parametrov.
	Delitev zapisov (split)	Omogoča delitev toka zapisov glede na izbrani pogoj.
	Združitev zapisov (union)	Omogoča združitev več tokov zapisov z enako strukturo.
	Uporaba skriptnega jezika	Omogoča realizacijo preoblikovanj podatkov z uporabo skriptnega jezika (Microsoft Visual Basic ali Microsoft C#)
	Počasi se spreminjajoče dimenzije (PSSD)	Omogoča avtomatsko vejitev zapisov glede na obstoječo vsebino dimenzijske tabele in izbrani tip PSSD na nivoju atributa.

Preglednica 11. Gradniki opravi v SSIS

9.3.3. Opravilo Polnjenje dimenzije Konto GK

V razdelkih, ki sledijo, bomo podrobneje predstavili posamezna opravila procesa ETL. Na sliki 26 so prikazani posamezni koraki opravila *Polnjenje dimenzije Konto GK*. Opravilo lahko glede na aktivnosti procesa ETL razdelimo na tri dele in sicer izbor, preoblikovanje in nalaganje oz. dostavo.



Slika 26. Opravilo Polnjenje dimenzije Konto GK

Opravilo se tipično začne z gradnikom Branje zapisov, kjer s poljubno SQL poizvedbo preberemo izbrane zapise iz tabel podatkovnega vira. V našem primeru smo uporabili sledečo SQL poizvedbo:

```
SELECT No_,Name,Debit_Credit FROM [CPK, d_d_$G_L Account] WHERE [Account Type] = 0
```

Za lažje razumevanje vsebine naslednjih korakov, v katerih se izvaja preoblikovanje podatkov, je za opravilo *Polnjenje dimenzije Konto GK* v preglednici 12 predstavljen *Pregled toka podatkov*, ki smo ga omenili v podpoglavju 9.2. Na slikah 27 in 28 pa so prikazane nastavitve posameznih gradnikov, uporabljenih v aktivnosti preoblikovanja podatkov.

Naziv atributa v podatkovnem viru	Naziv tabele v podatkovnem viru	Naziv atributa v dimenzijski tabeli Konto GK	Opis transformacije
Št	Konto GK	Št	- brez transformacije
Ime	Konto GK	Ime	- brez transformacije
Uspeh / stanje	Konto GK	Uspeh / stanje	- transformacija vrednosti iz 0 v »Uspeh« ter iz 1 v »Stanje«
Št	Konto GK	Razred 1	- izločitev prvega znaka iz atributa Št (slika 27)
Št	Konto GK	Razred 2	- izločitev prvih dveh znakov iz atributa Št (slika 27)
Št	Konto GK	Razred 3	- izločitev prvih treh znakov iz atributa Št (slika 27)
Št, Ime	Konto GK	Št. Ime	- združevanje atributa Št in Ime v nov atribut (slika 27) - pretvorba podatkovnega tipa v standardno dolžino niza (slika 28)

Preglednica 12. Pregled toka podatkov dimenzijske tabele Konto GK

Derived Column Name	Derived Column	Expression	Data Type	Length
Razred1	<add as new column>	SUBSTRING(No_,1,1)	Unicode string [DT_WSTR]	1
Razred2	<add as new column>	SUBSTRING(No_,1,2)	Unicode string [DT_WSTR]	2
Razred3	<add as new column>	SUBSTRING(No_,1,3)	Unicode string [DT_WSTR]	3
Št. Ime temp	<add as new column>	No_ + " " + Name	Unicode string [DT_WSTR]	51
Uspeh/Stanje	<add as new column>	Debit_Credit == 0 ? "Uspeh" : "Stanje"	Unicode string [DT_WSTR]	6

Slika 27. Nastavitve v koraku Določitev Razredov, Uspeh_Stanje, Št_ Ime

Input Column	Output Alias	Data Type	Length
Št. Ime temp	Št. Ime	Unicode string [DT_WSTR]	55
No_	Št.	Unicode string [DT_WSTR]	50
Name	Ime	Unicode string [DT_WSTR]	30

Slika 28. Nastavitve v koraku Pretvorba podatkovnih tipov

Podatke je po zaključenem preoblikovanju potrebno zapisati v dimenzijsko tabelo. Za realizacijo dostave potrebujemo funkcionalnost, ki nam bi glede na spremembe v podatkovnem viru in izbrane tipe PSSD omogočala pravilen pristop k spremembi zapisov v dimenzijski tabeli. Pod pravilnim pristopom imamo v mislih izbiro ene izmed treh možnosti:

- zapisa ni potrebno spremeniti
- spremeniti je potrebno vrednost posameznih polj v zapisu
- v dimenzijsko tabelo je potrebno dodati nov zapis

Opisano funkcionalnost je mogoče učinkovito realizirati z uporabo gradnika PSSD, ki omogoča samodejno vežitev zapisov glede na spremembe podatkovnega vira in določitev tipa PSSD. Gradnik podpira PSSD tipa 1 (prepisovanje) in tipa 2 (hramba zgodovine). Spremembe tipa 3 (nadomestna realnost) gradnik ne podpira, saj ta zahteva poseg v strukturo dimenzijske tabele.

Gradnik PSSD pozna dodaten tip PSSD, ki ga Kimball ni predvidel in je primeren za tiste attribute, kjer sprememba njihove vrednosti pomeni napako v podatkovnem viru. Tip PSSD, ki ga imenujemo *Fiksna vrednost*, omogoča dodatno kontrolo pravilnosti zapisov v podatkovnem viru.

Zaslonska posnetka določitve poslovnega ključa ter tipov PSSD sta prikazana na sliki 29. V primeru uporabe PSSD tipa 2 je potrebno dodatno določiti še datumski polji, ki določata začetek in konec veljavnosti zapisa.

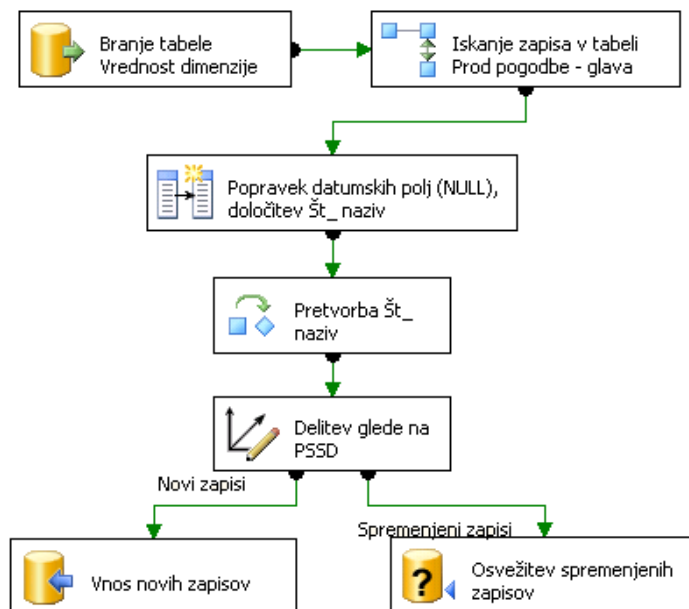
Input Columns	Dimension Columns	Key Type		Dimension Columns	Change Type
Ime	Ime	Not a key column	→	Ime	Changing attribute
Razred1	Razred1	Not a key column		Razred1	Changing attribute
Razred2	Razred2	Not a key column		Razred2	Changing attribute
Razred3	Razred3	Not a key column		Razred3	Changing attribute
Št.	Št.	Business key		Št. Ime	Changing attribute
Št. Ime	Št. Ime	Not a key column		Uspeh/Stanje	Fixed attribute
Uspeh/Stanje	Uspeh/Stanje	Not a key column			

Slika 29. Del nastavitve v koraku Delitev glede na PSSD

V zadnjih dveh korakih, ki se izvajata vzporedno, nad dimenzijsko tabelo izvedemo potrebne spremembe vrednosti atributov ali vnose novih zapisov.

9.3.4. Opravilo Polnjenje dimenzije SM

Ker so opravila polnjenja dimenzij *Spl. knjižna skupina proizvoda*, *Prilivi odlivi* ter *PC enostavna* in ne vsebujejo novih gradnikov, bomo njihov podrobni pregled preskočili. Na sliki 30 je prikazano opravilo polnjenje dimenzije *SM* (stroškovno mesto).



Slika 30. Opravilo Polnjenje dimenzije SM

V prvem koraku opravila iz tabele *Vrednost dimenzije* preberemo tiste zapise, ki ustrezajo dimenziji *SM*, kar storimo s sledečo poizvedbo:

```
SELECT [Dimension Code], Code, Name, Blocked FROM dbo.[CPK, d_d_$Dimension Value]
WHERE ([Dimension Code] = 'SM')
```

Za večino stroškovnih mest se v tabeli *Prod pogodbe – glava* nahajajo dodatni podatki. Za združitev atributov iz obeh tabel uporabimo gradnik *Iskanje zapisa*, kjer na podlagi atributa *Šifra* poiščemo pripadajoči zapis iz tabele *Prod pogodbe – glava*. Iskanje pripadajočega zapisa je realizirano s sledečo poizvedbo:

```
SELECT * FROM (SELECT * FROM dbo.[CPK, d_d_$Prod. Pogodbe - glava]) [RefTable]
WHERE [RefTable].[Št. SM] = [Št. SM]
```

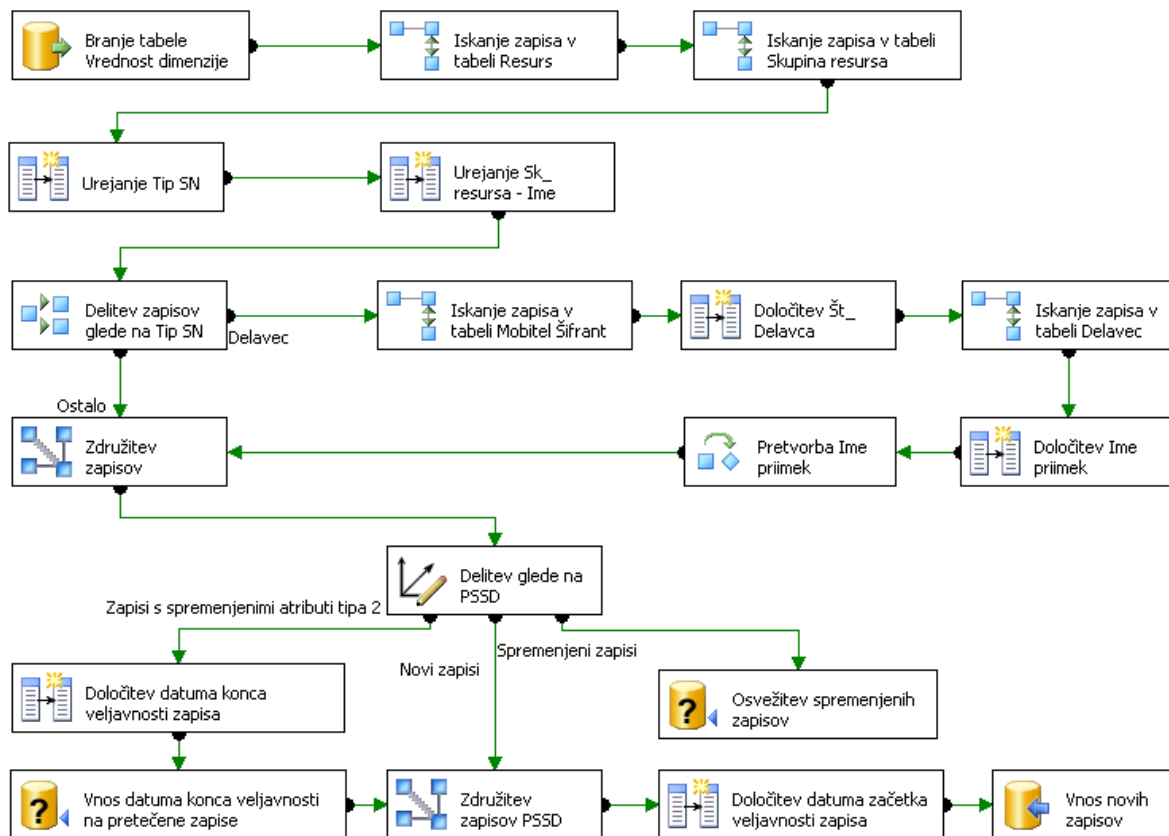
Gradnik *Iskanje zapisa* omogoča avtomatsko vejitev zapisov glede na njegovo uspešnost. V našem primeru te možnosti gradnika nismo uporabili, kar pomeni, da atributi iz iskane tabele za zapise, kjer iskanje ni uspešno, zavzamejo ničelne vrednosti (null value).

Po združitvi atributov obeh izvornih tabel se v tretjem koraku začne izvajati preoblikovanje podatkov. V korakih preoblikovanja popravimo vrednosti določenim atributom podatkovnega tipa datum ter iz atributov *Št* ter *Naziv* ustvarimo nov atribut *Št Naziv*. Opravilo *Polnjenje SM* se zaključi z določitvijo tipov PSSD ter dostavo podatkov v dimenzijsko tabelo *SM*.

Ker z izjemo omenjenih preoblikovanj vrednosti ostalih atributov v opravilu ostajajo nespremenjene, *Pregleda toka podatkov* ne bomo predstavili.

9.3.5. Opravilo Polnjenje dimenzije SN

Podatkovni vir za opravilo polnjenja dimenzije *SN* (stroškovni nosilec), ki je prikazano na sliki 31, predstavljajo zapisi kar petih tabel. Posamezen zapis v dimenzijski tabeli predstavlja stroškovni nosilec, ki lahko predstavlja stroj ali delavca, kar je razvidno iz atributa *Tip SN*.



Slika 31. Opravilo Polnjenje dimenzije SN

Ker je opravilo *Polnjenje SN* nekoliko obsežnejše, se ne bomo spustili v podrobnosti vsakega koraka, ampak le v dva koncepta, ki ju do sedaj še nismo omenili.

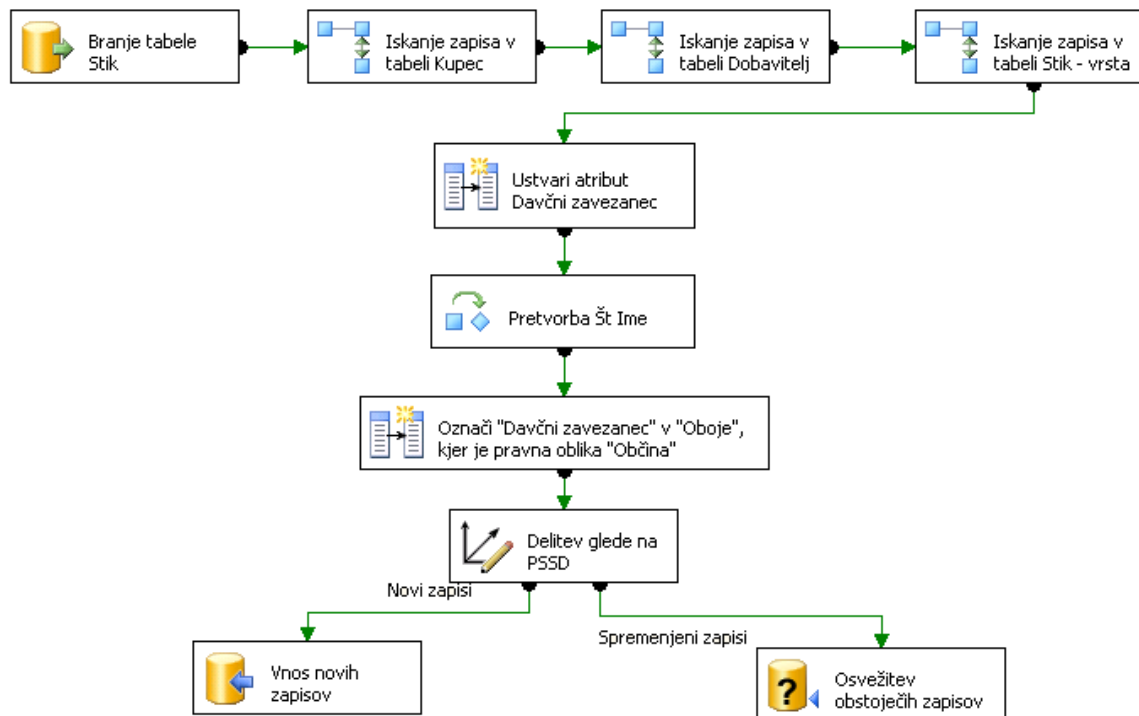
Prvi koncept je posledica dejstva, da so nekateri atributi v dimenziji namenjeni le določenim tipom SN. Iz tega razloga z namenom hitrejšega izvrševanja v koraku *Delitev zapisa glede na Tip SN* razdelimo potek opravlja. Nad zapisi, ki predstavljajo podatke o delavcih, izvedemo dodatne korake, s katerimi določimo vrednosti atributom, ki so zanje specifični, nato pa v koraku *Združitev zapisov* zapise ponovno združimo v enoten tok.

Realizacija drugega koncepta je nekoliko zahtevnejša in je posledica dejstva, da želimo pri dostavi atributa *Delavec – GSM omejitev* hraniti zgodovinske vrednosti. Tip 2 PSSD realiziramo tako, da ob vsaki spremembi vrednosti atributa:

- v obstoječi zapis zapišemo tekoči datum v atribut *Datum veljavnosti do*
- ustvarimo nov zapis, v katerem zapišemo tekoči datum v atribut *Datum veljavnosti od*

9.3.6. Opravilo Polnjenje dimenzij Stik

V dimenziji *Stik* hranimo podatke o poslovnih partnerjih združbe, ki so lahko kupci ali dobavitelji. Opravilo *Polnjenje dimenzije Stik* je predstavljeno na sliki 32.



Slika 32. Opravilo Polnjenje dimenzije Stik

Posebnost dimenzije *Stik* je atribut *Davčni zavezanec*, za katerega ne obstaja neposredni podatkovni vir in je izveden iz atributov *Davčna številka* in *Pravna oblika*. Atribut je zastavljen tako, da lahko zavzame vrednosti *zavezanec*, *nezavezanec* ali *oboje*. Poslovna logika za določitev vrednosti atributa je realizirana v procesu polnjenja s pomočjo dveh gradnikov *Preoblikovanje na nivoju zapisa*.

V prvem koraku atributu določimo vrednost glede na prvi znak atributa *Davčna številka*. V primeru, da prvi znak vsebuje črko, atributu določimo vrednost *zavezanec*, v nasprotnem primeru pa vrednost *nezavezanec*. Ozadje poslovne logike temelji na dejstvu, da se davčne številke zavezancev za DDV začnejo z oznako države (npr. SI), nezavezanci pa te predpone nimajo.

V prvem koraku vrednost atributa *Davčna številka* določimo s sledečim pogojnim izrazom:

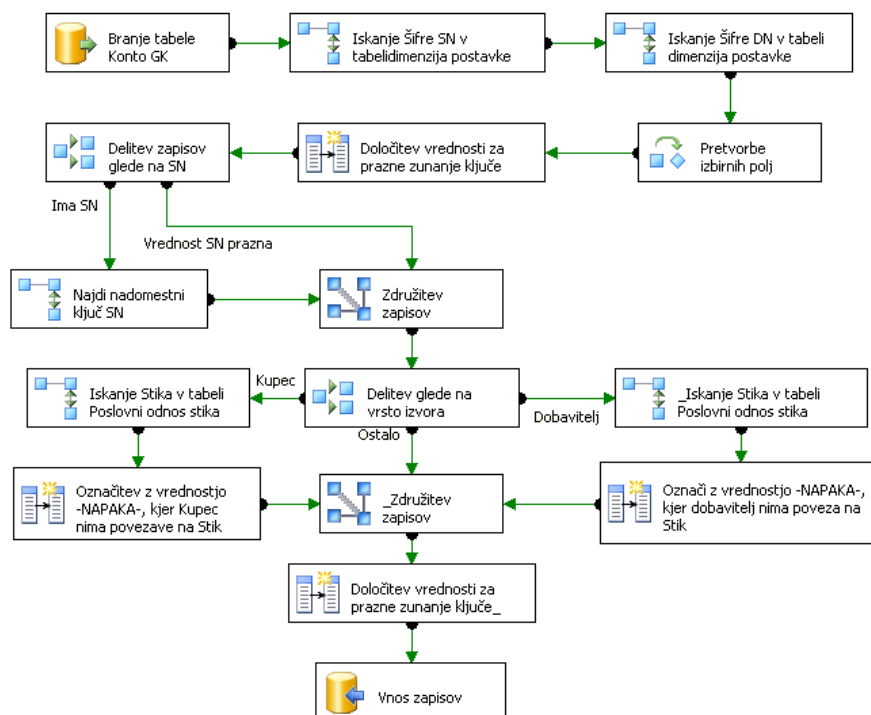
```
SUBSTRING([Davčna številka],1,1) >= "A" && SUBSTRING([Davčna številka],1,1) <= "Z"
? "Zavezanec" : "Nezavezanec"
```

Ker lahko občine, glede na vir financiranja, poslujejo kot davčni zavezanci ali nezavezanci, v drugem koraku stikom, ki predstavljajo občine, dodelimo vrednost *oboje*:

```
[Pravna oblika] == "občina" ? "Oboje" : [Davčni zavezanec]
```

9.3.7. Opravilo Polnjenje tabele dejstev Postavka GK

Za razliko od dimenzijskih tabel tabelo dejstev ob vsaki izvedbi procesa ETL v celoti zberemo in ponovno napolnimo. Za tak pristop smo se odločili, ker v podatkovnem viru nimamo učinkovitega mehanizma, s katerim bi lahko identificirali nove ali spremenjene zapise v izbranem časovnem obdobju. Kljub dejstvu, da je tak način polnjenja tabele dejstev časovno potraten, pa za branje, transformacijo in vnos približno 3,3 milijona zapisov, kolikor jih trenutno vsebuje tabela dejstev, potrebujemo manj kot 5 minut. Opravilo polnjenja tabele dejstev *Postavka GK* je prikazano na sliki 33.



Slika 33. Polnjenje tabele dejstev Postavka GK

V področnem podatkovnem skladišču *Glavna knjiga* so dimenzijske tabele s tabelo dejstev povezane preko primarnih (poslovnih) ključev. Izjema je dimenzijska tabela *SN*, ki je zaradi uporabe PSSD tipa 2 s tabelo dejstev povezana preko nadomestnega ključa.

Vrednost nadomestnega ključa določimo v koraku *Najdi nadomestni ključ SN*. To storimo tako, da v dimenzijski tabeli *SN* poiščemo najnovejši veljaven zapis, ki vsebuje pripadajočo vrednost poslovnega ključa:

```
SELECT * from (select * from [dbo].[SN II]) [refTable]
WHERE [refTable].[Veljavnost zapisa od] < [Datum knjiženja] AND [refTable].[Šifra SN] = [Šifra SN]
ORDER BY [Veljavnost zapisa od] DESC
```

V povezavi s polnjenjem tabele dejstev *Postavka GK* omenimo še določitev vrednosti zunanjega ključa *Stik* za primere, kjer za določeno šifro dobavitelja ali kupca ni pripadajočega zapisa v tabeli *Stik*. Ker gre v takih primerih za nekonsistentnost podatkovnega vira, zunanjemu ključu določimo vrednost *-NAPAKA-*.

9.4. Določitev fizične zasnove

Opis fizične zasnove področnega podatkovnega skladišča *Glavna knjiga* bomo razčlenili glede na opravila, ki so bila predstavljena v razdelku 7.4.2. Čeprav določitev fizične zasnove obravnavamo v okviru področnega podatkovnega skladišča *Glavna knjiga*, pa standardi za imenovanje objektov, določitve v povezavi z načrtovanjem fizičnega modela in nekatere tehnične lastnosti podatkovne baze veljajo za podatkovno skladišče v celoti.

9.4.1. Določitev standardov za imenovanje objektov

Določitev standardov smo v primerjavi z metodologijo življenjskega cikla nekoliko razširili. Standarde smo razdelili na naslednja področja:

- standardi imenovanja tabel in atributov
- standardi določitve vrednosti atributov podatkovnega tipa izbor (option)
- standardi določitve zunanjih ključev in degeneriranih dimenzij v tabeli dejstev
- standardi snovanja procesa ETL

Za imenovanje tabel in atributov v dimenzijskem podatkovnem modelu veljajo naslednja določila:

- imena tabel in atributov so navedena v slovenskem jeziku
- imena tabel in atributov, kjer je mogoče, sovpadajo z imenovanjem v ERP sistemu *Navision*
- v primeru, da je vsebina atributa glede na ime lahko za uporabnika zavajajoča, se ime razširi z dodatnim pojasnilom (npr. imenom tabele izvora)

V ERP sistemu *Navision* poznamo podatkovni tip *Izbor*, katerega atributi lahko zavzemajo tudi prazne vrednosti. V izogib napačnim razlagam s strani poslovnih uporabnikov, prazne vrednosti pretvorimo v:

- vrednost *nedefinirano*, kjer prazna vrednost pomeni, da v podatkovnem viru vrednost ni bila določena
- vrednost *ostalo*, kjer je izbor prazne vrednosti v podatkovnem viru namenski in je posledica tega, da želena vrednost ne sovпада z nobeno izbirno vrednostjo atributa

Pri določitvi vrednosti atributov v tabeli dejstev, ki predstavljajo zunanje ključe in degenerirane dimenzije veljajo naslednja določila:

- v primeru, da določen zapis v tabeli dejstev ni povezan z dimenzijsko tabelo, vrednosti zunanjega ključa za to dimenzijsko tabelo dodelimo ničelno vrednost
- v primeru, da je vrednost atributa, ki predstavlja degenerirano dimenzijo, prazen niz, le tega pretvorimo v ničelno vrednost
- za vrednosti zunanjih ključev uporabimo vrednosti ključev izvirnega sistema (poslovne ključe), razen za primere, ko dimenzijske tabele s tabelo dejstev povezujemo z uporabo nadomestnega ključa

Pri snovanju procesa ETL velja priporočilo, da že v koraku branja tabel izvirnega sistema imena atributov prevedemo v slovenščino.

9.4.2. Načrtovanje fizičnega modela

Fizični podatkovni model je razširitev logičnega podatkovnega modela, ki smo ga predstavili v podpoglavju 9.2. Ker želimo lastnosti fizičnega podatkovnega modela, kot so primarni ključi, dovoljenja za ničelne vrednosti in podatkovni tipi, v čim večji meri poenotiti na nivoju celotnega podatkovnega skladišča, smo za načrtovanje fizičnega podatkovnega modela določili nekaj enostavnih pravil.

V naših področnih podatkovnih skladiščih so dimenzijske tabele v večini primerov povezane s tabelo dejstev z uporabo poslovnega ključa. Vrednost ključa, ki predstavlja primarni ključ dimenzijske tabele, pridobimo neposredno iz podatkovnega vira. Kjer za povezavo dimenzijske tabele in tabele dejstev uporabimo nadomestni ključ, pa kot primarni ključ dimenzijske tabele določimo atribut podatkovnega tipa *Celo število*. Vrednost ključa se v teh primerih določi ob vnosu novega zapisa v dimenzijsko tabelo in sicer avtomatsko, na nivoju podatkovnega baze.

Ničelne vrednosti so v dimenzijskih tabelah dovoljene za attribute, ki v področnem podatkovnem skladišču niso del hierarhij in ne predstavljajo primarnega ključa dimenzijskih tabel. V tabelah dejstev so ničelne vrednosti dovoljene za degenerirane dimenzije in zunanje ključe, niso pa dovoljene za attribute, ki predstavljajo merljiva dejstva ali primarni ključ.

Izmed številnih podatkovnih tipov, ki jih omogoča podatkovna baza *SQL Server 2008* smo predpisali uporabo izključno naslednjih:

- za alfanumerične vrednosti: niz znakov (varchar) primerne dolžine s kodno tabelo WIN-1250
- za numerične vrednosti: decimalno število (decimal) z 20 mestno natančnostjo
- za datumska polja: datum (date)
- za nadomestne ključe: celo število (integer)

9.4.3. Načrtovanje agregatov in indeksov

Pri načrtovanju in realizaciji agregatov v SSAS je eno od ključnih vprašanj stopnja avtomatizacije, ki jo želimo prepustiti sistemu. Pri prvem izmed treh pristopov večino odločitev o tvorbi agregatov prepustimo sistemu, pri naprednejšem pa posamezne agregate določimo ročno. Tretji pristop omogoča določitev agregatov na podlagi uporabe sistema.

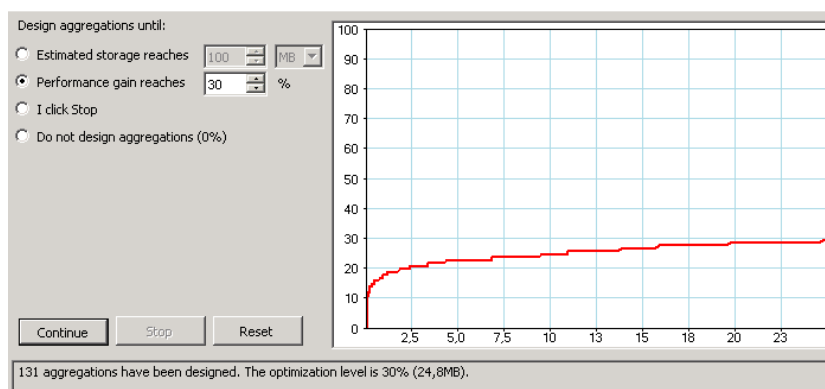
Pri avtomatiziranem pristopu, ki smo ga uporabili za naše podatkovno skladišče, začnemo načrtovanje agregatov z nastavitvijo posameznih atributov dimenzijskih tabel, pri čemer izbiramo med naslednjimi možnostmi:

- atribut naj bo uporabljen pri tvorbi agregatov, če predstavlja sestavni del vsaj ene hierarhije (unrestricted)
- vsak agregat naj vključuje izbrani atribut ali atribut iste dimenzije nižje na hierarhični lestvici (full)
- atribut naj ne bo vključen v tvorbo agregatov (none)

V primeru področnega podatkovnega skladišča *Glavna knjiga* smo pri vseh atributih izbrali prvo možnost. Predpostavljamo namreč, da bodo poslovni uporabniki v večini primerov uporabljali attribute, ki so vključeni v hierarhije.

V naslednjem koraku, ki je hkrati tudi zadnji, določimo število tvorjenih agregatov. To lahko storimo preko omejitve prostora, ki ga lahko agregati zasedejo v podatkovni bazi ali z določitvijo želene pridobitve hitrosti. V našem primeru smo sledili priporočilu, da je smiselna začetna nastavitev 20% do 30% povečanje hitrosti sistema [8].

Na sliki 34 prikazujemo zaslonski posnetek določitve agregatov za področno podatkovno skladišče *Glavna knjiga*. Graf na abscisi prikazuje potreben prostor za agregate, na ordinati pa predvideno povečanje hitrosti. Iz zaslonskega posnetka je razvidno, da je za ocenjeno 30% povečanje hitrosti tvorjeno 131 agregatov za katere je predvidena poraba 24,8 MB prostora.



Slika 34. Določitev agregatov za področno podatkovno skladišče Glavna knjiga

Kot alternativo ali nadgradnjo navedenemu pristopu omenimo še pristop, kjer se načrtovanja agregatov lotimo ročno, kar pomenim, da za vsak agregat sami izberemo želene attribute. Pristop je smiseln v primerih, ko želimo optimizirati delovanje določenih poročil.

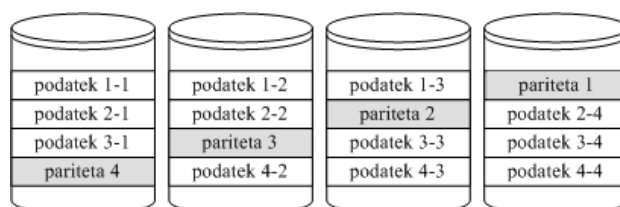
Tvorjenje indeksov v področnem podatkovnem skladišču zaradi uporabe tehnologije MOLAP ni potrebno, lahko pa z njihovo uporabo pospešimo delovanje procesa ETL. Z namenom hitrejšega iskanja nadomestnega ključa v dimenzijski tabeli *SN* smo v tabeli določili dodaten indeks, ki vključuje atributa *Šifra SN* in *Datum veljavnosti od*. Z indeksom smo čas, potreben za polnjenje tabele dejstev *Postavka GK*, skrajšali iz 10 minut in 46 sekund na 4 minute in 22 sekund.

9.4.4. Določitev tehničnih lastnosti podatkovne baze

Izmed tehničnih lastnosti podatkovne baze, naštetih v razdelku 7.4.2, bomo predstavili dve, ki najbolj vplivata na hitrost delovanja sistema. Gre za izbiro RAID konfiguracije, ki smo jo omenili že v podpoglavju 8.3. in delitev podatkovne baze na particije.

Izbira RAID konfiguracije vpliva tako na hitrost delovanja relacijske podatkovne baze, ki vsebuje tabele dimenzijskega podatkovnega modela, kot tudi na hitrost delovanja področnega podatkovnega skladišča, v katerem so podatki zapisani z uporabo tehnologije MOLAP. V naši konfiguraciji se namreč obe podatkovni strukturi nahajata na isti logični particiji.

Ker v našem sistemu čas izvajanja procesa ETL ni kritičen, želimo pa si čim hitrejšje izvajanje poizvedb nad področnimi podatkovnimi skladišči, stremimo k RAID konfiguraciji, ki zagotavlja čim hitrejšje branje podatkov. Iz tega razloga smo se odločili za konfiguracijo RAID 5, ki se izkaže kot dobra izbira za sisteme z več kot 70% bralnih operacij, med katere nedvomno lahko uvrstimo tudi podatkovna skladišča. Počasnejše pisanje konfiguracije RAID 5 v primerjavi z RAID 1 oziroma RAID 10 je posledica potrebe po izračunu paritet [21]. Shematski prikaz naše konfiguracije RAID 5, v kateri smo uporabili 4 fizične diske, je prikazan na sliki 35.



Slika 35. Shematski prikaz konfiguracije RAID 5

Kot razvidno iz slike 30, RAID 5 zagotavlja konsistentnost podatkov v primeru odpovedi katerega koli, a največ enega fizičnega diska.

Delitev podatkovne baze na particije bi lahko pomenila delitev podatkovne baze na več datotek in njihovo razpršitev na več logičnih diskovnih particij, kar je sorodno delovanju RAID polja. V našem primeru tega pristopa nismo uporabili. Pod omenjenim terminom si namreč razlagamo tehnologijo, omenjeno v razdelku 4.2.3., ki omogoča razdelitev podatkov MOLAP strukture na več particij (na isti logični diskovni particiji), s čimer na večprocesorskih sistemih omogočimo paralelno izvajanje poizvedb.

Uporaba particij je smiselna za tabele dejstev, ki vsebujejo veliko število zapisov. Particije ustvarimo tako, da zapise v tabeli dejstev razdelimo na podlagi izbranega kriterija. Pri delitvi je potrebno biti izredno pozoren, da v particijah zajamemo vse zapise tabele dejstev, ter da se noben izmed zapisov ne pojavlja v več particijah. Napaka te vrste bi namreč povzročila napačne končne rezultate.

V področnem podatkovnem skladišču *Glavna knjiga* smo dimenzijsko tabelo *Postavka GK* razdelili na dve particiji, za delitveni kriterij pa uporabili dimenzijo *Datum*. V prvi particiji se tako nahajajo zapisi do vključno leta 2007, v drugi pa od leta 2008 dalje. Na sliki 36 je prikazan zaslonski posnetek določitve particij področnega podatkovnega skladišča *Glavna knjiga*, kjer je razvidno, da za vsako particijo lahko izberemo lasten način agregiranja.

	Partition Name ▲	Source	Estimated Rows	Storage Mode	Aggregation Design
1	Postavka GK do vključno 2007	SELECT [dbo].[Postavka GK II].[Št. ...	2102275	MOLAP	Postavka GK
2	Postavka GK od vključno 2008	SELECT [dbo].[Postavka GK II].[Št. ...	1230791	MOLAP	Postavka GK

Slika 36. Particiji področnega podatkovnega skladišča Glavna knjiga

9.5. Snovanje kocke OLAP





Moderna orodja, kot je SSAS, pri snovanju kock OLAP omogočajo določitev lastnosti in funkcionalnosti, ki v metodologiji življenjskega cikla niso neposredno omenjena. Iz tega razloga smo v metodologijo dodali aktivnost, ki smo jo poimenovali *Snovanje kocke OLAP* in jo uvrstili v podatkovno sled, neposredno za aktivnost *Načrtovanje in realizacija procesa ETL*.

9.5.1. Merljiva dejstva

Snovanje kocke OLAP z uporabo SSAS smo začeli z določitvijo skupin merljivih dejstev, katerim podatkovni vir praviloma predstavlja osrednja tabela dejstev. V posebnih primerih, ko želimo med rezultati prikazati tudi attribute, ki se nahajajo izven tabele dejstev, pa lahko podatkovni vir zanje predstavlja tudi dimenzijska tabela. Znotraj posameznih skupin smo snovanje nadaljevali z določitvijo merljivih dejstev, ki se poslovnim uporabnikom v poročilih prikazujejo kot rezultati.

Za vsako merljivo dejstvo je potrebno izbrati atribut, ki predstavlja njegov podatkovni vir ter privzeti način agregiranja. Za aditivna dejstva je način agregiranja praviloma seštevanje (sum), pri pol aditivnih dejstvih pa je smiselna izbira način zadnji z vrednostjo (last non empty). V SSAS lahko izbiramo še med nekaj načini agregiranja, kot so na primer štetje (count), najmanjša vrednost (min), največja vrednost (max) in povprečje (average).

Na sliki 37 je preko posnetka zaslona prikazan seznam merljivih dejstev za kocko OLAP *Glavna knjiga*.

	Name	Measure Group	Data Type	Aggregation
	Promet	Postavka GK	Double	Sum
	V Breme	Postavka GK	Double	Sum
	V Dobro	Postavka GK	Double	Sum
	Mesečna kvota GSM	SN	Double	LastNonEmpty

Slika 37. Merljiva dejstva področnega podatkovnega skladišča Glavna knjiga

9.5.2. Uporaba dimenzij

V SSAS so dimenzije in skupine merljivih dejstev lahko povezane na različne načine. Povezave kocke OLAP *Glavna knjiga* so preko posnetka zaslona predstavljene na sliki 33.

Podobno, kot je veljalo za skupine merljivih dejstev, podatkovni vir dimenzij praviloma predstavljajo dimenzijske tabele, v posebnih primerih pa lahko vir predstavlja tudi tabela dejstev. Kot razvidno iz slike 38, sistem omogoča, da fizično ena tabela predstavlja podatkovni vir hkrati dimenziji kot tudi skupini merljivih dejstev.

Measure Groups		
Dimensions	Postavka GK	SN
Datum	Date	Postavka GK
PC	Šifra	
SM	Šifra	
Konto GK	Št	
Stik	Št	
Spl Knjižna Skupina Proizv...	Šifra	
Postavka GK	Št Postavke	
SN	Ključ SN	Ključ SN
Prilivi Odlivi	Šifra Prilivaodliva	

Slika 38. Povezanost dimenzij in skupin merljivih dejstev kocke OLAP Glavna knjiga

V SSAS lahko izbiramo med štirimi vrstami povezave med dimenzijami in skupinami merljivih dejstev:

- običajna povezava (regular)
- povezava dejstvo (fact)
- večkratna povezava (many-to-many)
- podatkovno rudarjenje (data mining)

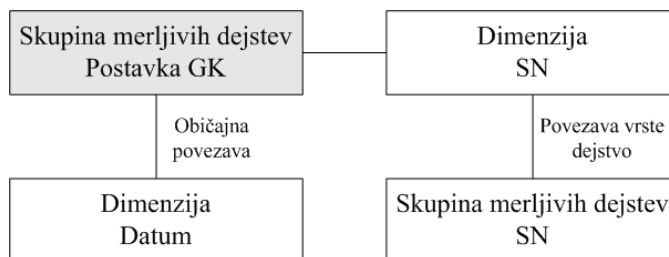
Za običajno vrsto povezave med dimenzijo in skupino merljivih dejstev se odločimo takrat, ko želimo realizirati klasično funkcionalnost dimenzije, to je, omejevanje rezultatov glede na različne vrednosti atributov, ki se nahajajo v dimenzijski tabeli. V našem primeru je običajna vrsta povezave uporabljena med skupino merljivih dejstev *Glavna knjiga* in dimenzijami *Datum*, *PC*, *SM*, *Konto GK*, *Stik*, *Spl. knjižna skupina proizvoda*, *SN* ter *Prilivi odlivi*.

Skupina merljivih dejstev *Glavna knjiga* poleg ključev za povezavo dimenzijskih tabel in merljivih dejstev vsebuje tudi attribute, ki jih želimo uporabiti za omejevanje rezultatov, a zanje nismo ustvarili dimenzijskih tabel. Take vrste atributov smo obravnavali v razdelku 4.1.1., in jih imenujemo degenerirane dimenzije. Za zagotovitev njihove funkcionalnosti je potrebno uporabiti povezavo vrste dejstvo. Gre za povezavo, pri kateri podatkovni vir tako za dimenzijo, kot tudi za skupino merljivih dejstev predstavlja ista fizična tabela (v našem primeru tabela dejstev *Postavka GK*).

Večkratno povezavo med dimenzijo in skupino merljivih dejstev določimo na osnovi obstoječe povezave vrste dejstvo. V primeru, ko želimo tako strukturo povezati z drugo dimenzijo, lahko to storimo preko vmesne skupine merljivih dejstev. Primer take uporabe v kocki OLAP *Glavna knjiga* je vezan na atribut *Delavec – GSM omejitev*, kjer sta skupina merljivih dejstev *SN* in dimenzija *SN* povezani s povezavo vrste dejstvo. Ker želimo omogočiti pravilen prikaz zgodovinskih vrednosti atributa *Delavec – GSM omejitev* med

rezultati poročil, je potrebno strukturo *SN* povezati z dimenzijo *Datum*, kar je mogoče preko vmesne skupine merljivih dejstev *Postavka GK*.

Na sliki 39 je na pravkar opisanem primeru prikazana večkratna povezava dimenzij s skupinami merljivih dejstev, na kateri je vmesna skupina merljivih dejstev obarvana s sivo barvo.



Slika 39. Večkratna povezava na primeru skupine merljivih dejstev SN

SSAS omogoča realizacijo modelov podatkovnega rudarjenja s podatki iz kocke OLAP. Z uporabo izbranega algoritma podatkovnega rudarjenja je mogoče odkrivati vzorce v podatkih ali napovedovati njihovo bodočo vrednost. Tako zbrane podatke nato za potrebe nadaljnjih analiz shranimo v dimenzije vrste podatkovno rudarjenje. V našem podatkovnem skladišču podatkovno rudarjenje trenutno ni implementirano.

9.5.3. Izračunana dejstva

Tabele dejstev vsebujejo zunanje ključe, degenerirane dimenzije in merljiva dejstva, katerih vrednosti so zapisane neposredno v podatkovni bazi, ki jo uporablja kocka OLAP. Za izračunana dejstva v SSAS navedeno ne velja, saj se njihova vrednost izračunava sproti na podlagi izraza, zapisanega v jeziku MDX.

Za lažje razumevanje logike, potrebne za implementacijo izračunanega dejstva Saldo je predhodno potrebno obrazložiti vsebino merljivega dejstva *Znesek* in koncept zapiranja izkaza poslovnega uspeha ob koncu leta.

Vrednost merljivega dejstva *Znesek* predstavlja nominalno vrednost posameznega zapisa v glavni knjigi oziroma knjižbe. V primeru, da gre za knjižbo v breme, je vrednost merljivega dejstva *Znesek* enaka merljivemu dejstvu *V breme*, v primeru knjižbe v dobro, pa vrednost merljivega dejstva *Znesek* predstavlja negativno vrednost merljivega dejstva *V dobro*. Ko zorni kot našega pogleda dvignemo z nivoja posameznega zapisa na nivo analiz, merljivo dejstvo *Znesek* dobi nekoliko drugačen pomen. Predstavlja namreč razliko med knjižbami v dobro in tistimi v breme v določenem časovnem obdobju. Ker želimo terminologijo podatkovnega skladišča čim bolj poenotiti s tisto v ERP sistemu *Navision* smo iz tega razloga za potrebe analiz merljivo dejstvo *Znesek* poimenovali *Promet*.

Posamezen konto glavne knjige je lahko označen kot konto izkaza uspeha ali bilance stanja. Konti izkaza uspeha (razredi 4,7 in 8) se po knjigovodskih pravilih konec leta »zaprejo«, kar pomeni, da je vrednost salda v začetku knjigovodskega leta enaka 0. Knjižbe zapiranje se v sistemu *Navision* evidentirajo na dodaten, umetno določen datum, ki je uvrščen med 31.

decembrom in 1. januarjem. Tak pristop omogoča pravilne izpise prometa, tudi ko izbrano obdobje vključuje zadnji dan tekočega leta.

Izračunano dejstvo *Saldo* predstavlja razliko med knjižbami v breme in dobro do zadnjega dne izbranega časovnega obdobja, razumemo ga lahko torej kot stanje na dan. Tako kot dejstvo *Promet*, želimo tudi izračunano dejstvo *Saldo* implementirati tako, da bo njegova vrednost usklajena z vrednostjo v sistemu *Navision*.

Izračun vrednosti merljivega dejstva *Saldo* je realiziran na podlagi naslednjih določil:

- knjižbe, ki predstavljajo zapiranje izkaza uspeha iz podatkovnega vira, ne prenašamo v področno podatkovno skladišče;
- za konte izkaza uspeha vrednost salda izračunamo tako, da seštejemo promet od začetka poslovnega leta do zadnjega dne izbranega obdobja;
- za konte bilance stanja izračunamo vrednost salda tako, da seštejemo celoten promet do zadnjega dne izbranega obdobja.

Opisan pristop nam omogoča hkraten prikaz pravilne vrednosti prometa in salda tako za konte bilance stanja kot tudi izkaza uspeha. Realizacija izračunanega dejstva *Saldo* v jeziku MDX je sledeča:

```
CREATE MEMBER CURRENTCUBE.[Measures].Saldo AS
Case
When [Konto GK].[Razred1] Is [Konto GK].[Razred1].&[4] Or
     [Konto GK].[Razred1] Is [Konto GK].[Razred1].&[7] Or
     [Konto GK].[Razred1] Is [Konto GK].[Razred1].&[8] Then
(
SUM (
PeriodsToDate(
[Datum].[Year - Half Year - Quarter - Month - Date].[Year],
[Datum].[Year - Half Year - Quarter - Month -
Date].CurrentMember), [Measures].[Promet])
)
else
(
SUM (
PeriodsToDate(
[Datum].[Year - Half Year - Quarter - Month - Date].[(All)],
[Datum].[Year - Half Year - Quarter - Month -
Date].CurrentMember), [Measures].[Promet])
)
)
```

9.5.4. Ključni kazalniki poslovanja

Tako kot izračunana dejstva tudi ključne kazalnike poslovanja izračunavamo sproti s pomočjo izrazov v jeziku MDX, pri čemer je pomembno, da za njihov izračun lahko uporabimo tako dejstva kot tudi izračunana dejstva.

Ključnim kazalnikom uspeha lahko določimo trende in ciljne vrednosti, katerih izpolnjevanje nato prikazujemo s pomočjo grafičnih označb. Za potrebe našega projekta smo se omejili na izračun vrednosti kazalnikov in določitev nominalnih ciljnih vrednosti.

Iz nabora temeljnih kazalnikov poslovanja slovenskega računovodskega standarda 29 – *računovodsko preučevanje* smo za realizacijo izbrali tiste kazalnike poslovanja, ki jih v združbi izračunavamo za potrebne letnega poročila.

Ker je način izračuna podoben za vse izbrane kazalnike, kot primer podrobno prikažimo le izračun kazalnika *Stopnja osnovnosti investiranja*, ki je definiran kot delež knjigovodske vrednosti osnovnih sredstev v sredstvih.

```
(
SUM([Konto GK].[Razred2].&[00],[Measures].[Saldo])+
SUM([Konto GK].[Razred2].&[02],[Measures].[Saldo])+
SUM([Konto GK].[Razred2].&[03],[Measures].[Saldo])+
SUM([Konto GK].[Razred2].&[04],[Measures].[Saldo])+
SUM([Konto GK].[Razred2].&[05],[Measures].[Saldo])+
SUM([Konto GK].[Razred3].&[130],[Measures].[Saldo])
)
/
(
SUM([Konto GK].[Razred1].&[0],[Measures].[Saldo])+
SUM([Konto GK].[Razred1].&[1],[Measures].[Saldo])+
SUM([Konto GK].[Razred1].&[3],[Measures].[Saldo])+
SUM([Konto GK].[Razred1].&[6],[Measures].[Saldo])
)
```

Za izračun kazalnikov je potrebno določiti razrede kontov glavne knjige, ki predstavljajo osnovne knjigovodske kategorije, kar prikazujemo v preglednici 13. Pri izračunu je za vse kategorije uporabljeno izračunano merljivo dejstvo *Saldo* z izjemo kategorije *Dividende poslovnega leta*, kjer je za izračun potrebno uporabiti merljivo dejstvo *V dobro*.

Oznaka	Knjigovodska kategorija	Skupine Kontov GK za izračun
OS	Osnovna sredstva (po knjigovodski vrednosti)	00+02+03+04+05+130
S	Sredstva	0+1+3+6
K	Kapital	4+7+810+813+90+91+92+930+95
OVS	Obveznosti do virov sredstev	2+4+7+810+813+90+91+92+930+95+96+97+98
K+DD+ PČR	Vsota kapitala in dolgoročnih dolgov (skupaj z rezervacijami) ter dolgoročnih pasivnih časovnih razmejitev	4+7+810+813+90+91+92+930+95+96+97+98
OS+ DAČR+ NN+ DFN+DPT	Vsota osnovnih sredstev in dolgoročnih aktivnih časovnih razmejitev (po knjigovodskih vrednosti), naložbenih nepremičnin, dolg. finančnih naložb in dolg. posl. terjatev	00+01+02+03+04+05+06+07+08+130
LS	Likvidna sredstva	10+11
KO	Kratkoročne obveznosti	2-29
LS+KT	Vsota likvidnih sredstev in kratkoročnih terjatev	10+11+12+133+134+139+14+15+16
KS	Kratkoročna sredstva	1+3-6-130-131-19
PP	Poslovni prihodki	76 + 79 - (Δ 63) - (Δ 600010)
PO	Poslovni odhodki	40+41+43+44+450+47+48+702+72+750
ČD	Čisti dobiček obračunskega obdobja	4+7+810+813
D	Dividende poslovnega leta	275020

Preglednica 13. Prikaz izračuna knjigovodskih kategorij

Enajst ključnih kazalnikov poslovanja, ki so po slovenskem računovodskem standardu razdeljeni v pet kategorij [26], smo realizirali kot količnik knjigovodskih kategorij, prikazanih v preglednici 11.

Temeljni kazalniki financiranja (vlaganja):

- stopnja lastniškosti financiranja (K / OVS)
- stopnja dolgoročnosti financiranja ($K + DD + PČR / OVS$)

Temeljni kazalniki stanja investiranja (naložbenja):

- stopnja dolgoročnosti investiranja ($OS + DAČR + NN + DFN + DPT / S$)
- stopnja osnovnosti investiranja (OS / S)

Temeljni kazalniki vodoravnega finančnega ustroja:

- koeficient kapitalske pokritosti osnovnih sredstev (K / OS)
- koeficient kratkoročne pokritosti kratkoročnih obveznosti (KS / KO)
- koeficient neposredne pokritosti kratkoročnih obveznosti (LS / KO)
- koeficient pospešene pokritosti kratkoročnih obveznosti ($LS + KT / KO$)

Temeljni kazalniki gospodarnosti:

- koeficient gospodarnosti poslovanja (PP / PO)

Temeljni kazalniki dobičkonosnosti:

- koeficient čiste dobičkonosnosti kapitala ($ČD / K$)
- koeficient dividendnosti osnovnega kapitala (D / K)

Za zaključek pregleda ključnih kazalnikov poslovanja si pogledjmo še realizacijo MDX izraza za izračun poslovnih prihodkov, ki se od ostalih razlikuje v tem, da je izračun potrebno določiti spremembo vrednosti ene izmed kategorij.

Spremembo zaloge (razred 63 in konto 600010) izračunamo tako, da vrednost zaloge predhodnega obdobja odštejemo od vrednosti zaloge tekočega obdobja.

```
SUM([Konto GK].[Razred2].&[76],[Measures].[Saldo])+
SUM([Konto GK].[Razred2].&[79],[Measures].[Saldo])-
(
SUM([Konto GK].[Razred2].&[63],[Measures].[Saldo])-
SUM([Datum].[Year - Half Year - Quarter - Month - Date].PREVMEMBER,[Konto
GK].[Razred2].&[63]),[Measures].[Saldo])
)
-
(
SUM([Konto GK].[Št].&[600010],[Measures].[Saldo])-
SUM([Datum].[Year - Half Year - Quarter - Month - Date].PREVMEMBER,[Konto
GK].[Št].&[600010]),[Measures].[Saldo])
)
```

9.6. Primeri uporabe

Nad kocko OLAP *Glavna knjiga* lahko z orodji poslovnega obveščanja, ki podpirajo izvajanje OLAP operacij, ustvarimo veliko število uporabnih poročil. Ker tudi orodje za splošno uporabo, kot je *Microsoft Excel*, podpira hierarhije oziroma vrtanje v globino, so poročila lahko interaktivna.

Uporabniki, ki imajo določena znanja o strukturi in vsebini podatkov, lahko poročila na enostaven način ustvarijo tudi sami. Take vrste poročil, ki so ustvarjena glede na trenutne potrebe, imenujemo ad-hoc poročila.

Za zaključek poglavja o področnem podatkovnem skladišču *Glavna knjiga* bomo preko zaslonskih posnetkov predstavili nekaj primerov poročil, kot jih vidijo poslovni uporabniki sistema.

Slika 40 prikazuje stroške storitev pri ustvarjanju proizvodov in opravljanju storitev (razred 410) za obdobje od leta 2006 do 2009 po prihodkovnih centrih združbe. Vrtanje v globino je omogočeno tako po dimenziji *Datum*, kot tudi dimenziji *PC*.

4	Promet	Oznake stolpcev				
5	Oznake vrstic	Calendar 2006	Calendar 2007	Calendar 2008	Calendar 2009	Skupna vsota
6	410010 Stroški storitev podizvajalcev	4.490.499,02	2.660.198,18	5.399.018,55	2.960.517,40	15.510.233,15
7	Asfaltna baza	36.090,63	268,80	1.250,00	181,98	37.791,41
8	Kamnomol Črni Kal	83,46	200,00		33.134,55	33.418,01
9	Kamnomol Razdrto	617,59	5.905,60	0,00	12.784,02	19.307,21
10	PC Gradnje	4.062.136,44	2.328.136,02	5.020.949,93	2.586.253,08	13.997.475,47
11	PC Mehanizacija	160,60		2.798,60	899,36	3.858,56
12	PC Vzdrževanje	363.050,99	313.795,29	373.228,98	317.749,61	1.367.824,87
13	Uprava	28.359,31	11.892,47	791,04	9.514,80	50.557,62
14	410011 Stroški dela - kooperanti	351.533,41	674.046,08	1.068.347,08	606.897,20	2.700.823,77
15	Kamnomol Razdrto		100,00			100,00
16	PC Gradnje	351.533,41	673.946,08	1.068.347,08	604.863,20	2.698.689,77
17	PC Mehanizacija				1.593,00	1.593,00
18	Uprava				441,00	441,00
19	410020 Miniranje	368.535,86	422.134,24	463.767,47	348.124,90	1.602.562,47
20	Kamnomol Črni Kal	160.475,53	149.504,95	197.743,33	188.550,63	696.274,44
21	Kamnomol Razdrto	208.060,33	272.629,29	266.024,14	149.832,54	896.546,30
22	Uprava				9.741,73	9.741,73
23	410030 Preiskava kamnitih materialov	15.435,35	10.448,18	16.612,38	10.146,23	52.642,14
24	Asfaltna baza	2.009,26	605,00	1.268,29		3.882,55
25	Kamnomol Črni Kal	1.013,82	1.140,48	2.195,94	2.080,00	6.430,24
26	Kamnomol Razdrto	3.734,09	6.608,51	4.863,86	4.356,14	19.562,60
27	PC Gradnje		-1.171,41	7.793,04	3.710,09	10.331,72
28	Uprava	8.678,18	3.265,60	491,25	0,00	12.435,03
29	410040 Druge proizvodne storitve	101.916,80	49.360,55	87.719,09	10.382,00	249.378,44
30	Asfaltna baza		8.910,00	264,00		9.174,00
31	Kamnomol Črni Kal	1.561,13		6.667,00	35,00	8.263,13
32	Kamnomol Razdrto	1.250,00	2.358,24	11.159,94		14.768,18
33	PC Gradnje	4.168,51	16.813,87	32.761,47	4.255,00	57.998,85
34	PC Vzdrževanje	73.701,42	5.866,13	24.818,33	5.073,00	109.458,88
35	Uprava	21.235,74	15.412,31	12.048,35	1.019,00	49.715,40
36	Skupna vsota	5.327.920,44	3.816.187,23	7.035.464,57	3.936.067,73	20.115.639,97

Slika 40. Stroški storitev pri ustvarjanju proizvodov in opravljanju storitev

Na sliki 41 prikazujemo deleže prihodkov združbe po prihodkovnih centrih.

4	Promet	Oznake stolpcev				
5	Oznake vrstic	Calendar 2006	Calendar 2007	Calendar 2008	Calendar 2009	Skupna vsota
6	Asfaltna baza	3,54%	3,88%	2,08%	2,36%	2,86%
7	Kamnlom Črni Kal	4,70%	5,34%	3,13%	3,24%	3,98%
8	Kamnlom Razdrto	3,55%	6,56%	3,82%	2,39%	3,98%
9	PC Gradnje	60,76%	57,99%	69,89%	62,79%	63,58%
10	PC Mehanizacija	0,71%	0,25%	0,23%	0,13%	0,32%
11	PC Vzdrževanje	23,08%	25,42%	20,67%	28,65%	24,15%
12	Uprava	3,67%	0,55%	0,19%	0,45%	1,13%
13	Skupna vsota	100,00%	100,00%	100,00%	100,00%	100,00%

Slika 41. Prihodki po prihodkovnih centrih

Na sliki 42 so predstavljeni temeljni kazalniki poslovanja združbe, realizirani z uporabo ključnih dejavnikov poslovanja (KPI).

2	Vrednosti	Calendar 2006	Calendar 2007	Calendar 2008	Calendar 2009
3	Koeficient čiste dobičkonosnosti kapitala	0,10	0,04	0,05	0,01
4	Koeficient dividendnosti osnovnega kapitala	0,05	0,07	0,07	
5	Koeficient gospodarnosti poslovanja	1,06	1,06	1,05	1,05
6	Stopnja dolgoročnosti financiranja	0,49	0,43	0,37	0,43
7	Stopnja lastniškosti financiranja	0,31	0,27	0,22	0,24
8	Stopnja dolgoročnosti investiranja	0,61	0,58	0,61	0,62
9	Stopnja osnovnosti investiranja	0,49	0,48	0,53	0,54
10	Koeficient kapitalske pokritosti OS	0,64	0,56	0,42	0,44
11	Koeficient kratkoročne pokritosti kratkoročnih obveznosti	0,76	0,71	0,55	0,58
12	Koeficient neposredne pokritosti kratkoročnih obveznosti	0,02	0,01	0,00	0,00
13	Koeficient pospešene pokritosti kratkoročnih obveznosti	0,53	0,33	0,35	0,36

Slika 42. Temeljni kazalniki poslovanja

Slika 43 prikazuje prilive in odlive združbe po vrstah za izbrani mesec.

4	Promet	Oznake stolpcev		
5	Oznake vrstic	Odliv	Priliv	Skupna vsota
6	financiranje	-1.396.712,57	475.056,59	-921.655,98
7	KRAJK. POSOJILA IN DEPOZITI		475.000,00	475.000,00
8	OBRESTI BANK (PREJETE)		56,59	56,59
9	OBRESTI BANK IN POVEZANI STROŠKI	-81.712,57		-81.712,57
10	VRAČILA DOLG.POSOJIL	-30.000,00		-30.000,00
11	VRAČILA KRAJK. POSOJIL IN DEPOZITOV	-1.285.000,00		-1.285.000,00
12	investiranje	-79.657,68		-79.657,68
13	DOBAVITELJI - NABAVA OSN. SREDSTEV	-2.445,30		-2.445,30
14	LEASINGI	-77.212,38		-77.212,38
15	poslovanje	-1.551.335,92	2.554.094,91	1.002.758,99
16	DAVEK OD DOHODKOV PRAVNIH OSEB	-2.057,84		-2.057,84
17	DOBAVITELJI - SPLOŠNI STROŠKI	-36.558,30		-36.558,30
18	DOBAVITELJI - SPREMENLJIVI STROŠKI	-497.037,96		-497.037,96
19	DRUGI ODLIVI	-408.243,98		-408.243,98
20	DRUGI PRILIVI		139.839,57	139.839,57
21	KUPCI		1.995.096,17	1.995.096,17
22	OBRESTI DOBAVITELJEV	-13,41		-13,41
23	OBRESTI KUPCI (ZO)		245,00	245,00
24	PLAČE IN DRUGI OSEBNI PREJ. ZAPOS.	-607.274,43		-607.274,43
25	POLOG GOTOVINE		742,00	742,00
26	ŽE UPOŠTEVANI ODLIVI	-150,00		-150,00
27	ŽE UPOŠTEVANI PRILIVI		418.172,17	418.172,17
28	Skupna vsota	-3.027.706,17	3.029.151,50	1.445,33

Slika 43. Prilivi in odlivi združbe v izbranem mesecu

10. Preostala področna podatkovna skladišča

Preostalih področnih podatkovnih skladišč ne bomo predstavili tako celovito in natančno, kot je to veljalo za področno podatkovno skladišče *Glavna knjiga*. Omejili se bomo namreč zgolj na predstavitev specifičnih rešitev, ki smo jih uporabili pri realizaciji in so lahko posledica posebnosti v izvornih podatkih ali strmenja k zagotavljanju zahtevanih funkcionalnosti. Poleg posebnosti pri realizaciji bomo z namenom boljšega razumevanja vsebine posameznih področnih podatkovnih skladišč predstavili še dimenzijski podatkovni model in nekaj primerov uporabe.

V nadaljevanju poglavja bomo predstavili področna podatkovna skladišča *Saldakonti*, *Materialno poslovanje* in *Evidenca dela*.

10.1. Področno podatkovno skladišče Saldakonti

Področno podatkovno skladišče *Saldakonti* pokriva knjigovodski področji saldakontov kupcev ter dobaviteljev oziroma procesa prodaje in nabave. Z namenom učinkovitejše spremljave podatkov o poslovnih partnerjih, ki z združbo sodelujejo tako v vlogi kupca kot dobavitelja so podatki o kupcih in dobaviteljih združeni v enotno dimenzijsko tabelo *Stik*.

Področno podatkovno skladišče *Saldakonti* poslovnim uporabnikom omogoča izvajanje podrobnih analiz terjatev in obveznosti na dan ali v obdobju glede na zapadlost in vrsto terjatve oz. obveznosti.

10.1.1. Dimenzijski podatkovni model

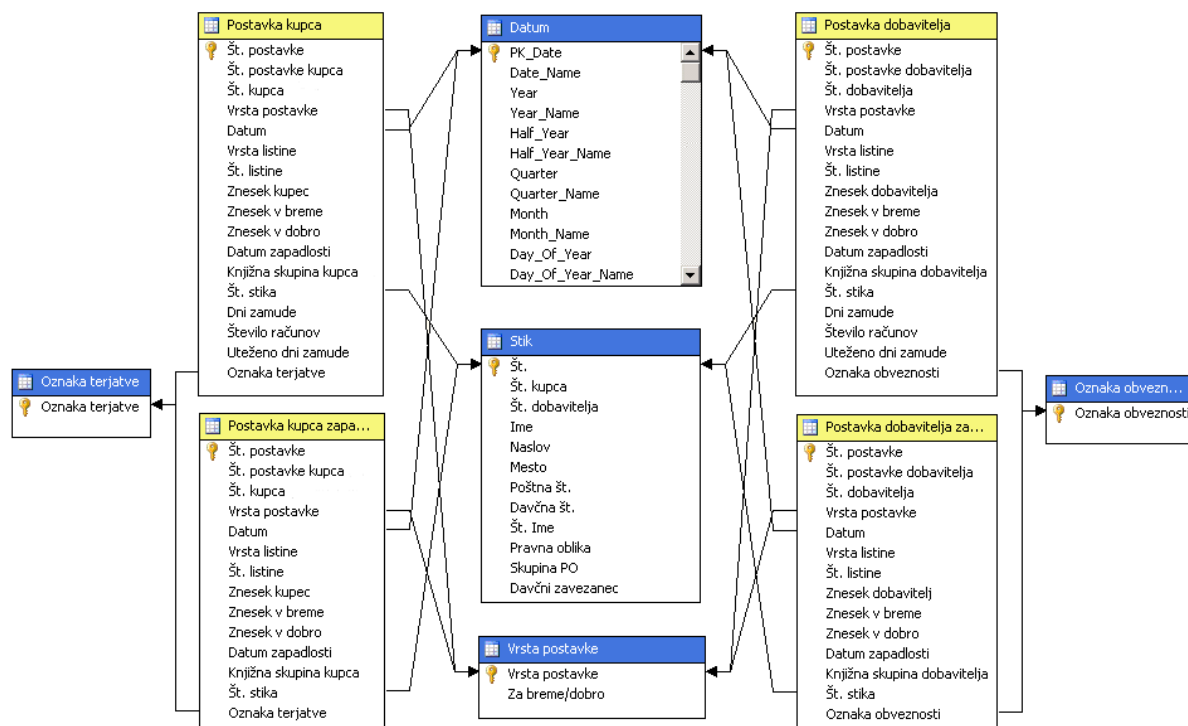
Posebnost dimenzijskega modela področnega podatkovnega skladišča *Saldakonti*, ki ga prikazujemo na sliki 44, je v tem, da vsebuje kar štiri, z rumeno barvo označene tabele dejstev:

- postavka kupca
- postavka kupca zapadlost
- postavka dobavitelja
- postavka dobavitelja zapadlost

Vzrok za ločeni tabeli dejstev postavka kupca in dobavitelja je v tem, da ima vsaka od tabel dejstev sorodne, a vsebinsko različne merljive vrednosti. Dimenzijski model bi lahko zasnovali tudi z enotno tabelo dejstev, ki bi vsebovala tako postavke kupca kot tudi dobavitelja, a bi posledično vsebovala podvojeno število merljivih vrednosti. Zaradi enostavne uporabe več tabel dejstev v SSAS se za to možnost nismo odločili.

Za zagotovitev podatkov o zapadlih terjativah oz. obveznostih smo zasnovali tabeli dejstev *Postavka kupca zapadlost* in *Postavka dobavitelja zapadlost*, katerih vsebino predstavljajo nekoliko preoblikovani zapisi dimenzijskih tabel *Postavka kupca* oz. *Postavka dobavitelja*.

Zrno vseh štirih tabel dejstev predstavlja poslovni dogodek v povezavi s kupcem oz. dobaviteljem, kot je na primer izstavitve računa, dobropisa ali realizacija plačila.



Slika 44. Dimenzijski podatkovni model področnega podatkovnega skladišča Saldakonti

10.1.2. Predstavitev specifičnih rešitev

Posebnost področnega podatkovnega skladišča *Saldakonti* je, kot že rečeno, uporaba štirih tabel dejstev, ki ob snovanju kocke OLAP predstavljajo štiri skupine merljivih dejstev. Ker gre pri kupcih in dobaviteljih za enako strukturo podatkov in način realizacije se bomo v nadaljevanju predstavitev osredotočili le na kupce, torej na tabeli dejstev *Postavka kupca* in *Postavka kupca zapadlost*.

Za lažje razumevanje realizacije preoblikovanj podatkov iz tabele dejstev *Postavka kupca* v *Postavka kupca zapadlost* na primeru najprej predstavimo izračunano merljivo dejstvo *K saldo*, ki izvira iz tabele dejstev *Postavka kupca*.

Predpostavimo, da se v tabeli dejstev *Postavka kupca* nahajata zgolj dva zapisa, ki sta prikazana na sliki 45.

Datum knjiženja	Datum zapadlosti ...	Vrsta postavke	Vrsta listine	Št. listine	Št. kupca	Znesek
15.11.07	15.12.07	Zač. postavka	Račun	IR-07-05223	975	3.890,28
14.01.08	15.12.07	Zapiranje	Plačilo	K01-11	975	-3.890,28

Slika 45. Primer zapisov iz tabele dejstev Postavka kupca

Vrednost merljivega dejstva *K saldo* se izračunava na dan kot seštevek atributa *Znesek kupec* za vse zapise do zadnjega dneva izbranega obdobja. Gre torej za enak način izračuna, kot pri merljivem dejstvu *Saldo* v področnem podatkovnem skladišču *Glavna knjiga*. Vsebinsko *K Saldo* predstavlja višino terjatev na določen datum.

Merljivo dejstvo *K zapadli saldo* vsebinsko predstavlja višino zapadlih terjatev. Če želimo končnemu uporabniku omogočiti, da z omejitvijo dimenzije *Datum*, ki je vezana na atribut *Datum knjiženja*, dobi hkraten podatek tako za saldo kot tudi zapadli saldo je potrebno zapise v tabeli dejstev *Postavka kupca zapadlost*, ki v osnovi predstavljajo kopijo zapisov tabele dejstev *Postavka kupca*, nekoliko preoblikovati.

Preoblikovanje je realizirano tako, da v primeru, ko vrednost atributa *Datum zapadlosti* predstavlja kasnejši datum od vrednosti atributa *Datum knjiženja*, slednjemu dodelimo vrednost atributa *Datum zapadlosti*.

Na sliki 46 sta prikazana zapisa v tabeli dejstev *Postavka kupca zapadlost*, ki sovpadata z zapisi iz slike 45.

Datum knjiženja	Datum zapadlosti ...	Vrsta postavke	Vrsta listine	Št. listine	Št. kupca	Znesek
15.12.07	← 15.12.07	Zač. postavka	Račun	IR-07-05223	975	3.890,28
14.01.08	15.12.07	Zapiranje	Plačilo	K01-11	975	-3.890,28

Slika 46. Primer zapisov iz tabele dejstev *Postavka kupca zapadlost*

Primer izračunanih vrednosti merljivih dejstev *K saldo* in *K zapadli saldo*, ki se nanaša na zapise iz slik 45 in 46 za izbrana časovna obdobja je prikazan v preglednici 14.

Na dan	K saldo	K zapadli saldo
do 14.11.07	0,00	0,00
od 15.11.07 do 14.12.07	3.890,28	0,00
od 15.12.07 do 13.1.08	3.890,28	3.890,28
od 14.1.08	0,00	0,00

Preglednica 14. Primer izračuna vrednosti merljivega dejstva *K saldo* in *K zapadli saldo*

Z uporabo dimenzije *Datum* in merljivimi dejstvi *K Saldo* in *K zapadli saldo* lahko uporabnik torej pridobi podatke o vrednosti terjatev in zapadlih terjatev na dan. Za prikaz podatka o starostni strukturi terjatev, pa je potrebno uporabiti dimenzijo *Datum zapadlost*, ki si z dimenzijo *Datum* deli dimenzijsko tabelo. Dimenzija *Datum zapadlost* je s tabelami dejstev povezana preko istoimenskega atributa *Datum zapadlosti*, kar je razvidno iz slike 47.

Measure Groups				
Dimensions	Postavka Kupca	Postavka Kupca Zapadlost	Postavka Dobavitelja	Postavka Dobavitelja Zapad...
Datum	Date	Date	Date	Date
Stik	Št	Št	Št	Št
Postavka Kupca	Št Postavke			
Oznaka Terjatve	Oznaka Terjatve	Oznaka Terjatve		
Datum zapadlost	Date		Date	
Oznaka Obveznosti			Oznaka Obveznosti	Oznaka Obveznosti
Vrsta Postavke	Vrsta Postavke	Vrsta Postavke	Vrsta Postavke	Vrsta Postavke

Slika 47. Povezanost dimenzij in skupin merljivih dejstev področnega podatkovnega skladišča *Saldakonti*

10.1.3. Primeri uporabe

Na sliki 48 je prikazan delež zapadlih obveznosti v vseh obveznostih na zadnji dan izbranih mesecev.

3	D Delež zapadlosti	Oznake stolpcev ▾		
4	Oznake vrstic	October 2009	November 2009	December 2009
5	pravna oseba	0,26	0,30	0,17
6	odvetnik, notar	0,00	0,03	0,00
7	d.o.o.	0,24	0,25	0,08
8	s.p.	0,52	0,53	0,35
9	d.d.	0,21	0,34	0,43
10	društvo, zavod, klub	0,12	0,19	0,50
11	k.d.	1,00	1,00	1,00
12	Skupna vsota	0,26	0,30	0,17

Slika 48. Delež zapadlih obveznosti

Terjatve po vrstah na zadnji dan poslovnih let v obdobju od leta 2006 do 2009 so prikazane na sliki 49.

3	K Saldo	Oznake stolpcev ▾			
4	Oznake vrstic	Calendar 2006	Calendar 2007	Calendar 2008	Calendar 2009
5	Čista	4.694.690,78	5.110.913,02	7.489.903,61	6.591.073,25
6	Dvomljiva	142.936,08	215.241,31	272.422,43	278.417,39
7	Odvetnik	36.334,79	40.219,08	47.956,21	43.069,09
8	Tožena	125.429,05	126.836,90	211.631,85	209.807,52
9	Skupna vsota	4.999.390,70	5.493.210,31	8.021.914,10	7.122.367,25

Slika 49. Terjatve po vrstah

Na sliki 50 so razvidni podatki o številu računov, izdanih v posameznih mesecih, njihova skupna vrednost ter povprečna zamuda pri njihovem plačilu.

3	Vrednosti			
4	Oznake vrstic	<div><div></div>K Število računov</div>	K Znesek računov	K Povp. dni zamude
5	January 2009	293	2.285.974,14	36,12
6	February 2009	376	1.744.500,98	37,49
7	March 2009	564	2.678.369,58	30,92
8	April 2009	575	2.034.248,50	27,13
9	May 2009	598	2.465.873,44	21,56
10	June 2009	648	2.127.169,18	15,79
11	July 2009	599	2.954.032,71	30,71
12	August 2009	511	2.190.826,17	17,08
13	September 2009	550	2.308.051,24	15,47
14	October 2009	618	3.278.131,00	13,91
15	November 2009	482	2.395.938,30	17,60
16	December 2009	376	2.562.302,68	19,33
17	Skupna vsota	6.190	29.025.417,92	22,75

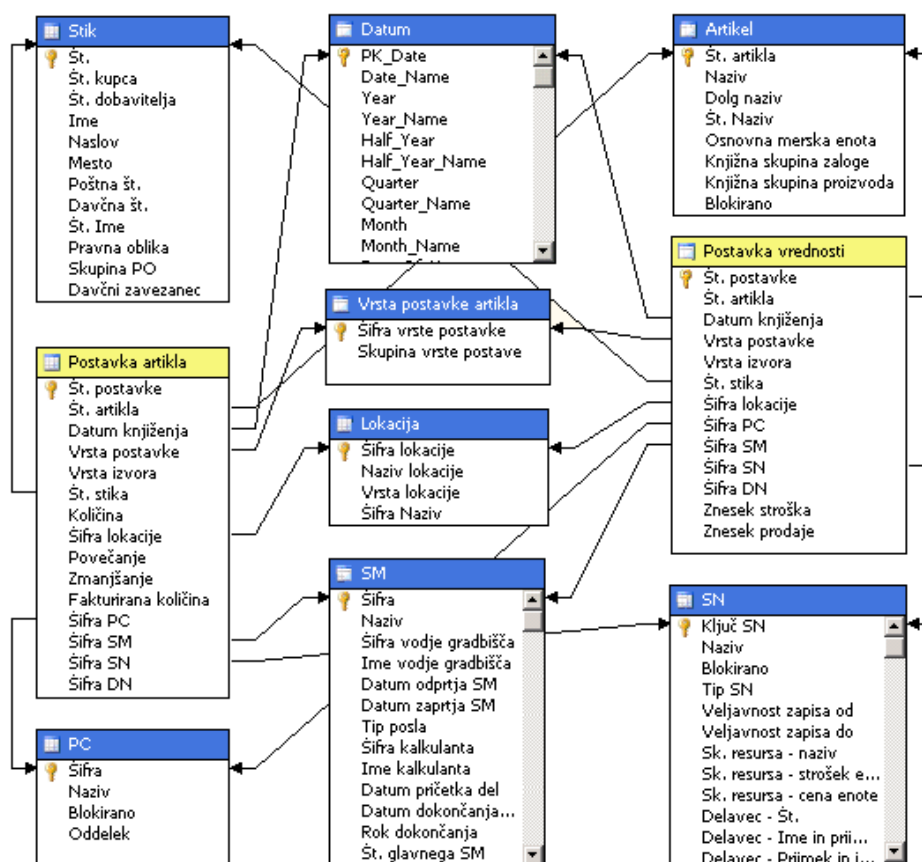
Slika 50. Analiza izdanih računov po mesecih

10.2. Področno podatkovno skladišče Materialno poslovanje

Ključna dimenzijska tabela področnega podatkovnega skladišča *Materialno poslovanje* je dimenzijska tabela *Artikel*, ki predstavlja šifrant materiala, drobnega inventarja in izdelkov združbe. Področno podatkovno skladišče *Materialno poslovanje* omogoča izvajanje količinskih in vrednostnih analiz o gibanju materiala in drobnega inventarja v skladiščih ter prodaji izdelkov med katere štejemo kamnolomske agregate in asfaltno maso.

10.2.1. Dimenzijski podatkovni model

Na sliki 51 je prikazan dimenzijski podatkovni model področnega podatkovnega skladišča *Materialno poslovanje*, na katerem sta tabeli dejstev označeni z rumeno barvo.



Slika 51. Dimenzijski podatkovni model področnega podatkovnega skladišča Materialno poslovanje

10.2.2. Predstavitev specifičnih rešitev

Posebnost področnega podatkovnega skladišča *Materialno poslovanje* predstavlja uporaba dveh tabel dejstev z različnima zrnoma, ki pa sta z dimenzijskimi tabelami povezani na enak način. Tabela dejstev *Postavka vrednosti* predstavlja osnovo za vrednostno analizo, *Postavka artikla* pa za analizo količin.

V tabeli dejstev *Postavka artikla* je zrno določeno kot posamezna transakcija nad artiklom, kot sta na primer prejem v skladišče ali izdaja iz skladišča. Ker pa so skladiščne transakcije lahko v začetku ovrednotene le okvirno, natančno pa se na podlagi regulacije ovrednotijo šele naknadno, zrno tabele dejstev *Postavka vrednosti* predstavlja posamezno ovrednotenje postavke artikla.

Poudariti velja, da končni uporabnik uporabe dveh tabel dejstev z različnima zrnoma v ničemer ne občuti kot pomanjkljivost.

10.2.3. Primeri uporabe

Na sliki 52 je prikazana količinska in vrednostna primerjava prodaje izdelkov po prihodkovnih centrih. Ker je prodaja visoko korelirana s sezonskim obdobjem se primerjava izvaja nad primerljivimi obdobji, v našem primeru nad prvimi četrtletji preteklih let.

5	Oznake stolpcev						
6	Asfaltna baza	Kamnomol Črni Kal		Kamnomol Razdrto			
7	Oznake vrstic	Prodajna količina (M3)	Znesek Prodaje	Prodajna količina (M3)	Znesek Prodaje	Prodajna količina (M3)	Znesek Prodaje
8	Quarter 1, 2007	9.714	389.800,16	34.704	314.528,02	31.465	219.258,79
9	Quarter 1, 2008	10.327	423.001,35	29.853	265.705,68	59.533	459.093,34
10	Quarter 1, 2009	9.485	504.323,97	38.674	276.857,27	23.635	197.017,47
11	Quarter 1, 2010	12.142	534.726,01	16.362	160.306,46	18.799	152.387,88
12	Skupna vsota	41.668	1.851.851,49	119.593	1.017.397,43	133.432	1.027.757,48

Slika 52. Prodaja izdelkov po prihodkovnih centrih

Slika 53 prikazuje stroške izdanih tonerjev in kartuš za obdobje petih preteklih četrtletij.

4	Strošek zmanjšanja	Oznake stolpcev					
5	Oznake vrstic	Quarter 1, 2009	Quarter 2, 2009	Quarter 3, 2009	Quarter 4, 2009	Quarter 1, 2010	Skupna vsota
6	901678 TONER HP Q2612A (KOS)	-1.364,68	-842,07	-1.224,00	-1.020,00	-660,26	-5.111,01
7	902433 TONER SAMSUNG ML-2010D3 (KOS)	-328,62	-647,84	-600,00	-792,00	-564,04	-2.932,50
8	902356 RICOH TONER TYPE FX200 (KOS)	-206,94	-544,51	-725,00	-580,00	-290,00	-2.346,45
9	901680 TONER HP Q5949A (KOS)	-278,23	-311,80	-156,00	-156,00	-52,00	-954,03
10	902415 BOBEN ENOTA ZA SPC210SF (KOS)	-406,03		-480,00			-886,03
11	902017 TONER RICOH CYAN 888315 (KOS)	-313,26		-251,75	-210,00		-775,01
12	902016 TONER RICOH MAGENTA 888314 (KOS)	-313,26				-461,75	-775,01
13	902550 TONER RICOH BLACK 402887 (KOS)	-200,69			-200,69	-332,82	-734,20
14	902295 Toner Ricoh TYPE 140 BLACK (KOS)	-138,78	-290,00	-300,00			-728,78
15	902018 TONER RICOH YELLOW 888313 (KOS)	-313,26		-251,75	-84,00		-649,01
16	902019 TONER RICOH BLACK 888312 (KOS)		-110,46		-378,64	-84,00	-573,10
17	902838 BOBEN ENOTA ZA 410/411 BARVNA (KOS)			-456,00			-456,00
18	901745 KARTUŠA HP 51645 (KOS)	-53,04	-111,49	-87,67	-120,00	-46,40	-418,60
19	901717 KARTUŠA HP 6578DE COLOR (KOS)	-140,10		-59,94	-98,65	-50,64	-349,33
20	901852 TONER HP C8061X (KOS)	-79,29	-123,58		-121,19		-324,06
21	902297 Toner Ricoh TYPE 140 MAGENTA (KOS)				-289,88		-289,88
22	901886 TONER CANON FX 10 (KOS)	-41,73	-114,76	-31,30	-46,00	-46,00	-279,79
23	902296 Toner Ricoh TYPE 140 CYAN (KOS)			-126,67	-140,00		-266,67
24	901744 TONER HP C7115A (KOS)	-53,18	-45,00	-137,38			-235,56
25	902414 CANON TONER C-EXV14 (KOS)				-120,11	-81,90	-202,01
26	902241 Toner CANON C-EXV12 (KOS)				-82,12	-91,25	-173,37
27	902128 CANON TONER C-EXV3 (KOS)	-91,80	-44,00	-33,00			-168,80
28	902898 TONER SAMSUNG MLT-D1052L (KOS)					-167,65	-167,65
29	902298 Toner Ricoh TYPE 140 YELLOW (KOS)	-160,03					-160,03
30	901905 KARTUŠA CANON PGI-5Bk (KOS)	-53,90	-25,20	-15,06	-27,06	-12,00	-133,22
31	901714 TONER SAMSUNG ML-1610D2 (KOS)		-115,67				-115,67
32	Skupna vsota	-4.536,82	-3.326,38	-4.935,52	-4.466,34	-2.940,71	-20.205,77

Slika 53. Stroški tonerjev in kartuš

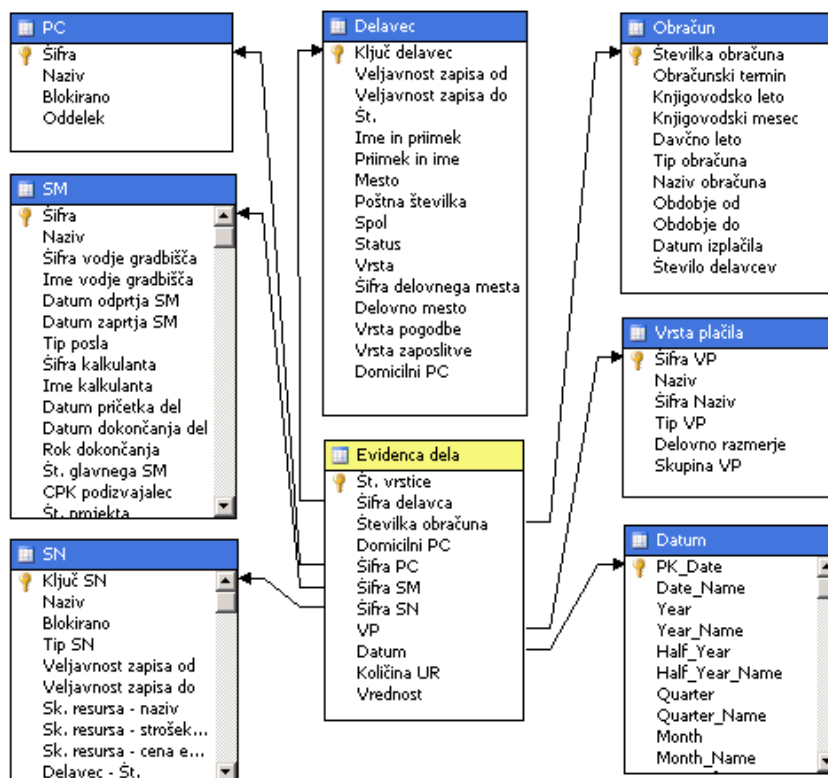
10.3. Področno podatkovno skladišče Evidenca dela

Zrno tabele dejstev *Evidenca dela* predstavlja posamezen zapis v istoimenski evidenci ERP sistema *Navision*, ki je osnova za obračun plač. Posamezen zapis evidence vsebuje število opravljenih ur delavca v določenem obdobju za izbrano vrsto plačila. V primeru, da izbrana vrsta plačila označuje povračilo stroškov (prehrana, kilometrina), se v zapisu poleg količine evidentira tudi vrednost predvidenega stroška. V področnem podatkovnem skladišču *Evidenca dela* je torej vrednostne analize mogoče izvajati le v povezavi s povračili stroškov.

Področno podatkovno skladišča *Evidenca dela* omogoča izvajanje analiz o opravljenih urah in predvidenih stroških povračil delavcev po prihodkovnih centrih, stroškovnih mestih in stroškovnih nosilcih. Preko dimenzijske tabele *Vrsta plačila* je mogoče izvajati analize v povezavi s povračili stroškov, odsotnostjo, nadurnim delom ipd. V primeru, da bi se izkazala potreba po vrednostnem analiziranju podatkov tudi izven okvira povračil stroškov, bi bilo tabelo dejstev smiselno razširiti z vpeljavo dodatnih atributov, kot na primer *Bruto*, *Bruto 2*, *Neto*. Dodatnim atributom bi podatkovni vir predstavljal modul *Plače* ERP sistema *Navision*, v katerem se ob obračunu plače, glede na količinski vnos iz modula *Evidenca dela*, ovrednotijo postavke vseh vrst plačila.

10.3.1. Dimenzijski podatkovni model

Na sliki 54 je prikazan dimenzijski podatkovni model področnega podatkovnega skladišča *Evidenca dela*.



Slika 54. Dimenzijski podatkovni model področnega podatkovnega skladišča *Evidenca dela*

10.3.2. Predstavitev specifičnih rešitev

V področnem podatkovnem skladišču *Evidenca dela* ni bilo potrebno poseči po specifičnih rešitvah ne v procesu ETL, kot tudi ne v dimenzijskem podatkovnem modelu. Pri snovanju kocke OLAP velja omeniti le delno aditivno izračunano dejstvo *Število delavcev*, ki se izračunava kot število različnih vrednosti atributa *Šifra delavca*. Vpeljava izračunanega dejstva *Število delavcev* je smiselna, ker se v poročilih pogosto izkaže potreba po podatku o številu delavcev, katerih evidence so bile uporabljene za izračun ostalih merljivih dejstev.

10.3.3. Primeri uporabe

Na sliki 55 je prikazana analiza odsotnosti v breme delodajalca po prihodkovnih centrih, v kateri prikazujemo število ur odsotnosti in število delavcev, ki so bili odsotni v določenem obdobju.

6		⊕ Asfaltna baza		⊕ Kamnolom Črni K		⊕ Kamnolom Razdrt		⊕ PC Gradnje		⊕ PC Mehanizacija		⊕ PC Vzdrževanje		⊕ Uprava	
7	Oznake vrstic	☑ UR	delavcev	UR	delavcev	UR	delavcev	UR	delavcev	UR	delavcev	UR	delavcev	UR	delavcev
8	June 2009		96	1	80	1	8	1	320	5	168	6	184	7	156
9	July 2009		32	1			56	1	640	12	136	4	280	10	32
10	August 2009		48	1	8	1	120	1	440	8	385	6	64	4	148
11	September 2009						56	2	416	11	504	6	88	4	184
12	October 2009		8	1			24	2	128	6	240	9	248	7	256
13	November 2009		80	1			8	1	424	9	488	9	536	8	128
14	December 2009				8	1	8	1	416	8	368	9	496	8	296
15	January 2010		152	1	32	1	24	1	536	10	536	10	328	6	144
16	February 2010		88	1			96	3	568	10	528	11	768	12	160
17	March 2010				32	1	80	2	248	11	464	8	408	11	196
18	April 2010		16	1			8	1	176	6	280	5	160	7	584
19	May 2010				56	1	32	1	104	4	80	4	128	6	348
20	Skupna vsota		520	5	216	3	520	5	4416	45	4177	39	3688	39	2632

Slika 55. Odsotnost v breme delodajalca v obdobju preteklih 12 mesecev

Slika 56 prikazuje število ur koriščenja letnega dopusta po mesecih leta 2009 za celotno združbo.

5		Vrednosti	
6	Oznake vrstic	Količina UR	Število delavcev
7	January 2009	4792	135
8	February 2009	3416	141
9	March 2009	3080	124
10	April 2009	4936	209
11	May 2009	3546	146
12	June 2009	7012	274
13	July 2009	7628	173
14	August 2009	10164	201
15	September 2009	4008	142
16	October 2009	3768	150
17	November 2009	3096	145
18	December 2009	8824	244
19	Skupna vsota	64270	311

Slika 56. Koriščenje letnega dopusta po mesecih leta 2009

11. Namestitev sistema

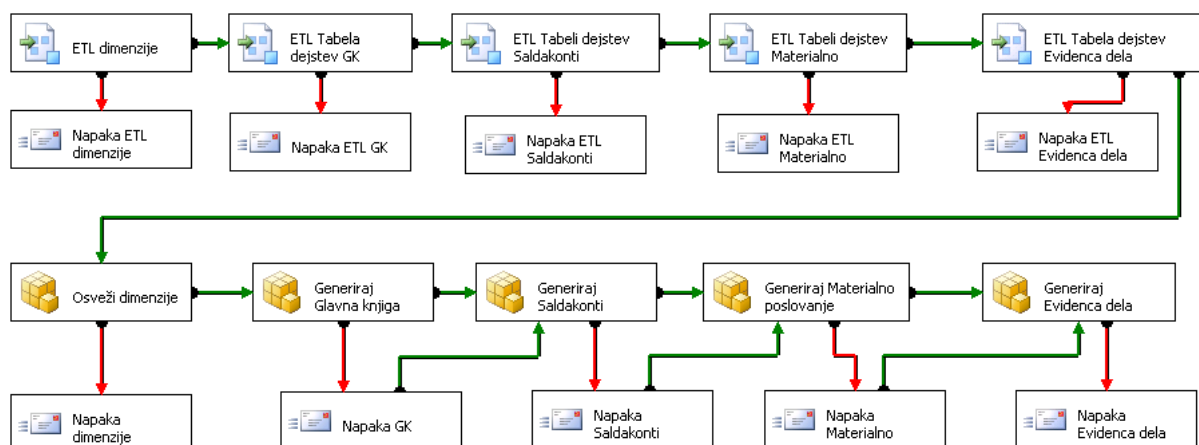
Kljub temu, da metodologija *Življenjskega cikla* v aktivnosti namestitve sistema obravnava le opravila, povezana z namestitvijo čelnega sistema, bomo v podpoglavju, ki sledi, najprej obravnavali namestitev zalednega sistema.

11.1. Namestitev zalednega sistema

Namestitev zalednega sistema predstavljata dve opravili, ki ju je potrebno izvesti za zagotovitev samodejnega osveževanja podatkovnega skladišča:

- določitev SSIS paketov, ki zagotavljajo osveževanje podatkovnega skladišča
- namestitev razporejevalnika

Za naše potrebe smo zasnovali globalen SSIS paket, ki združuje SSIS pakete ETL procesov nižjega nivoja, s čimer je zagotovljeno osveževanje celotnega podatkovnega skladišča. Dodatni funkcionalnosti globalnega SSIS paketa, ki je prikazan na sliki 57, predstavljata še samodejno obveščanje v primeru napak, ki je realizirano z uporabo elektronske pošte in osveževanje kock OLAP. Slednje se izvaja po uspešno zaključenih procesih ETL.



Slika 57. Globalni SSIS paket

Posamezna opravila so v SSIS paketih povezana s povezavami različnega tipa, ki določajo potek izvajanja. Po uspešno izvedenih opravilih se izvajanje globalnega SSIS paketa nadaljuje po povezavah zelene barve, v primeru napake pa po rdečih povezavah. Kot je razvidno iz slike, se v primeru napake v procesu ETL izvajanje paketa ustavi, v primeru napake pri osveževanju posameznih kock OLAP, pa se izvajanje nadaljuje.

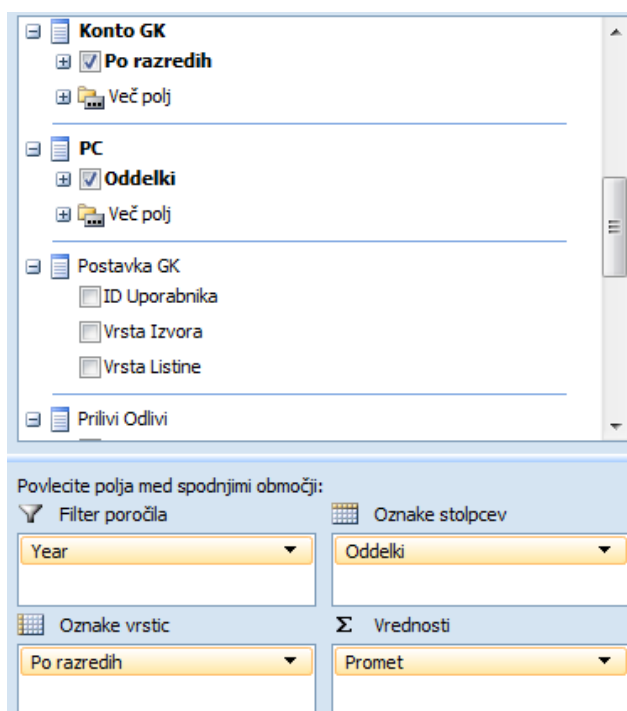
Zaradi uporabe modernega orodja za snovanje in obvladovanje procesa ETL in generiranja kock OLAP smo se odločili, da shrambe vmesnih podatkov ne bomo uporabili.

Ker smo osveževanje vseh področnih podatkovnih skladišč zasnovali z enim SSIS paketom, je namestitev razporejevalnika, ki skrbi za njegove periodične zagone, dokaj enostavna. Za naše potrebe smo določili izvajanje globalnega SSIS paketa pred začetkom vsakega delovnega dne.

11.2. Namestitev čelnega sistema

V začetnem obdobju delovanja sistema poslovnega obveščanja, kot že omenjeno, kot orodje poslovnega obveščanja uporabljamo *Microsoft Excel*. Iz tega razloga na računalnikih končnih uporabnikov ni potrebno nameščati dodatnih aplikacij.

Izobraževanje končnih uporabnikov se izvaja individualno, za kar smo se odločili zaradi različnega predhodnega znanja uporabnikov in različnih potreb pri uporabi. Izobraževanje sestoji iz predstavitve uporabniškega vmesnika, ki je na primeru področnega podatkovnega skladišča *Glavna knjiga* prikazan na sliki 58 in obrazložitve vsebine podatkov. Pred namestitvijo sistema je potrebno poskrbeti še za določitev pravic uporabnikov nad področnimi podatkovnimi skladišči v celoti ali natančneje nad posameznimi dimenzijami in tabelami dejstev.



Slika 58. Uporabniški vmesnik orodja za poslovno obveščanje

Kot je razvidno iz slike 58, je uporabniški vmesnik intuitiven, a vseeno zahteva določeno stopnjo razumevanja o področju, ki ga pokriva posamezno področno podatkovno skladišče. Napredni poslovni uporabniki ustvarijo novo poročilo tako, da izberejo želena merljiva dejstva, na osi vrtilne tabele prenesejo izbrane dimenzijske attribute in določijo omejitve. Poročilo lahko nato shranijo v obliki *Excelovega* dokumenta, ga preoblikujejo in kasneje osvežujejo z novimi podatki iz podatkovnega skladišča.

Preverba pravilnosti rezultatov se je izvajala že med snovanjem in testiranjem delovanja posameznih področnih podatkovnih skladišč. Ker pa področja uporabe sistema *Navision* in sistema poslovnega obveščanja za posamezne poslovne uporabnike v veliki meri sovpadajo, je sprotno preverjanje pravilnosti prepuščeno končnim uporabnikom.

12. Vzdrževanje in nadgradnje

V okviru vzdrževanja sistema poslovnega obveščanja se držimo priporočil metodologije *Življenjskega cikla*, ki v tem sklopu predvideva izvajanje opravil za zagotavljanje podpore in izobraževanja uporabnikov, učinkovitega delovanja in promoviranja sistema ter spremljavo uporabe.

Nadgradnje sistema smo za potrebe našega projekta razdelili na tri vsebinska področja:

- posodabljanje obstoječih področnih podatkovnih skladišč
- snovanje novih področnih podatkovnih skladišč
- nadgradnje ERP sistema z namenom zagotavljanja višje kakovosti podatkov
- uvedba naprednega orodja poslovnega obveščanja

V okviru posodabljanja obstoječih področnih podatkovnih skladišč bomo na podlagi lastnih ugotovitev in predlogov poslovnih uporabnikov po potrebi izvajali:

- dodajanje oziroma dopolnjevanje dimenzij, dimenzijskih atributov in hierarhij
- dodajanje merljivih vrednosti
- dodajanje izračunanih vrednosti in ključnih dejavnikov uspeha

S snovanjem novih področnih podatkovnih skladišč bomo nadaljevali, ko bodo obstoječa podatkovna skladišča dosegla določen nivo zrelosti. Slednje bomo ugotavljali predvsem na podlagi odzivov poslovnih uporabnikov med izvajanjem aktivnosti vzdrževanja, kot so na primer:

- obseg predlogov za izboljšave
- zadovoljstvo z obstoječim sistemom
- ocena obsega uporabe sistema

Na podlagi ugotovljenih napak v podatkih bomo z namenom izboljšanja njihove kakovosti skušali dopolnjevati ERP sistem na način, ki preprečuje njihovo ponovitev. Napake bomo, kot že do sedaj, skušali v čim večjem obsegu odpraviti na njihovem izvoru.

V roku šestih mesecev bomo postopoma prešli na uporabo naprednega orodja poslovnega obveščanja. Trenutno se nagibamo k produktom *NovaView* združbe *Panorama*, ki je specializirano za orodja poslovnega obveščanja. Uporabniki so po raziskavi združbe *Gartner Group* produkte *NovaView* uvrstili na prvo mesto med orodij za izvajanje OLAP operacij. Njihove pglavitne prednosti so predvsem v enostavnosti uporabe, ki se kaže v višjem odstotku izvajanju kompleksnih analiz s strani poslovnih uporabnikov v primerjavi s konkurenčnimi orodij ter integracija s paketom *Microsoft Office* [7]. Z uporabo omenjenih produktov želimo uporabnikom preko različnih kategorij orodij poslovnega omogočiti čim učinkovitejšo uporabniško izkušnjo, pri čemer bo osrednjo vlogo igrala uporaba nadzornih plošč.

Za zaključek omenimo še možnost bodoče nadgradnje sistema z orodji za poslovno analiziranje, česar pa v bližnji prihodnosti ne predvidevamo.

13. Sklep

Ključni rezultat magistrskega dela predstavlja realiziran sistem poslovnega obveščanja. Za doseg tega cilja je bilo potrebno izvesti številne aktivnosti, med katerimi lahko kot najpomembnejše izpostavimo:

- določitev tehnološke zasnove in platforme
- pripravo dimenzijskih podatkovnih modelov
- zagotavljanje kakovosti izvornih podatkov
- načrtovanje in realizacijo procesov ETL
- načrtovanje in realizacijo poročil
- snovanje kock OLAP

Med rezultate magistrskega dela, ki niso neposredno povezani z realizacijo sistema poslovnega obveščanja, prištevamo še predlog razširitve metodologije *Življenjskega cikla* z uvedbo dodatne aktivnosti, imenovane *Snovanje kock OLAP*.

Domneva, da nam med realizacijo sistema poslovnega obveščanja nekakovostni izvorni podatki ne bodo povzročali večjih težav, se je izkazala kot resnična. Napake, ki so bile odkrite v aktivnosti realizacije procesa ETL in pri uporabi sistema, so bile uspešno odpravljene na nivoju podatkovnega vira. Kjer je bilo možno, so bile v ERP sistemu *Navision* realizirane dodelave, ki onemogočajo ponovno pojavljanje napak iste vrste.

Med potekom realizacije sistema smo prišli do trdnega spoznanja, da je zaradi kompleksnosti in obsežnosti sistemov poslovnega obveščanja posamezno rešitev možno realizirati na mnogo različnih načinov. Prav iskanje optimalnih rešitev, ki so zahtevale poglobitev v podrobnosti Microsoftove platforme *SQL Server 2008*, je razvoj časovno najbolj obremenilo.

Sistem poslovnega obveščanja je potrebno neprestano dopolnjevati, njegove funkcionalnosti pa širiti. Ključnega pomena je torej, da realizacijo sistema poslovnega obveščanja razumemo kot proces. Sprotno prilagajanje sistema na spremembe poslovnih procesov in nove potrebe poslovnih uporabnikov je potreben pogoj za njegovo dolgoročno uspešnost.

S predvidenim preходом na tronivojsko arhitekturo bomo do uporabnikov lahko pristopili bolj individualno ter tako ponudili njim primernejši način dostopa do želenih informacij, kar bo postopoma povečevalo krog poslovnih uporabnikov. Informacije bodo uporabnikom na voljo v vsakem trenutku, saj bo dostop do sistema omogočen preko spletnega brskalnika. Kljub številnim izzivom, ki jih predvidevamo pri realizaciji omenjenega prehoda, bo potrebno neprestano zagotavljati visoko stopnjo pravilnosti rezultatov, ter tako ohranjati kredibilnost sistema. Ko bo večina aktivnosti, povezanih z vzpostavitvijo področnih podatkovnih skladišč in preходом na tronivojsko arhitekturo zaključenih, bo potrebno razmisliti o vzpostavitvi sistema za merjenje njegove uporabe. Povratne informacije, pridobljene iz omenjenega sistema, bi lahko igrale pomembno vlogo za prihodnje nadgradnje in zagotavljanje podpore poslovnim uporabnikom.

Po namestitvi prvih področnih podatkovnih skladišč smo prišli do ugotovitve, da poslovne odločitve, ki izvirajo iz informacij, pridobljenih iz sistema poslovnega obveščanja, predstavljajo pravo merilo kakovosti sistema poslovnega obveščanja. Kljub dejstvu, da je

posledice poslovnih odločitev v praksi izredno težko meriti, pa nas ugotovitev opominja na njihovo pomembnost.

V zadnjih nekaj letih je v povezavi s sistemi poslovnega obveščanja in odločanjem, kot posledice njihove uporabe, zaznati miselni preskok, ki bo v sistemih bližnje prihodnosti najverjetneje sprožil pomembne konceptualne premike. Z namenom pospeševanja kakovostnih poslovnih odločitev bo potrebno tesneje povezati informacije in odločitve. Na lestvici, ki sega od transakcijskih do strateških odločitev, bo potrebno zvišati nivo, do koder so informacije sistemov poslovnega obveščanja za odločitve ključnega pomena.

Za potrditev naše teze citirajmo besede raziskovalnega analitika združbe AMR Johna Haggertya, ki pravi [4]: »Sistemi poslovnega obveščanja so se do sedaj osredotočali predvsem na nadzorne plošče in posredovanje individualnih informacij poslovnim uporabnikom. Kar naenkrat pa je pozornost preusmerjena na integracijo sistemov poslovnega obveščanja znotraj poslovnih aplikacij. Informacije se vedno bolj povezuje v kontekst konkretnih poslovnih odločitev.«

Viri

- [1] M. Biere, *Business Intelligence for the Enterprise*, IBM Press, 2003, pogl. 3.
- [2] E. Codd, *A Relational Model of Data for Large Shared Data Banks*, Communications of the ACM, 1970.
- [3] (2002) P. Dasgupta. Killer Applications. Dostopno na:
<http://cactus.eas.asu.edu/partha/Columns/2002/07-01-killer-app.htm>
- [4] (2008) T. Davenport. 10 Principles of the New Business Intelligence. Harvard Business Publishing. Dostopno na:
http://blogs.harvardbusiness.org/davenport/2008/12/10_principles_of_the_new_busin.html
- [5] T. Davenport, J. G. Harris, *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, 2007.
- [6] W. Eckerson, *Performance Dashboards*, Wiley Inc, 2006.
- [7] J. Feiman, N. MacDonald, »Magic Quadrant for Business Intelligence Platforms«, Gartner RAS Core Research Note G00173700, 2010.
- [8] S. Harinath, R. Zare, S. Meenakshisundaram, M. Carroll, D. G. Lee, *Professional Microsoft SQL Server Analysis Services 2008 with MDX*, Wiley, 2009.
- [9] S. Harwood, *ERP: The Implementation Cycle*, Butterworth-Heinemann, 2003, pogl. 2.
- [10] (2004) C. Imhoff, N. Gelemmco, J. Geiger. Comparing the two data warehouse methodologies. Database and Network journal.
Dostopno na: http://findarticles.com/p/articles/mi_hb3234/is_3_34/ai_n29101628/
- [11] B. Inmon, *Bulding the data warehouse (3rd edition)*, Wiley Inc, 2002, pogl. 2.
- [12] R. Kimball, J. Caserta, *The Data Warehouse ETL Toolkit*, Wiley Inc, 2004.
- [13] R. Kimball, L. Reeves, M. Ross, W. Thornthwaite, *The Data Warehouse Lifecycle Toolkit*, Wiley Inc, 1998.
- [14] R. Kimball, M. Ross, *The Data Warehouse Toolkit (Second Edition)*, Wiley Inc, 2002.
- [15] J. Paredes, *The multidimensional data modeling Toolkit*, OLAP World Press, 2009, pogl. 2.
- [16] (2007) N. Raden. Business Intelligence 2.0: Simpler, More Accessible, Inevitable. Intelligent Enterprise.
Dostopno na: <http://www.intelligententerprise.com/showArticle.jhtml?articleID=197002610>

- [17] (2009) R. Sherman. Data Integration: All together now. Information management magazine.
Dostopno na: http://www.information-management.com/issues/2007_56/10014940-1.html
- [18] (2008) H. Stevens, C. Pettey. Gartner EXP Worldwide Survey. Gartner Press Releases.
Dostopno na: <http://www.gartner.com/it/page.jsp?id=587309>
- [19] R. J. Thierauf, *Effective business intelligence systems*, Greenwood Publishing Group, 2001, pogl. 4.
- [20] E. Turban, *Information technology for management: transforming organizations in the digital economy*, Wiley Inc, 2006, stran 505.
- [21] N. Vengurlekar, M. Vallath, R. Long, *Oracle Automatic Storage Management: Under the-Hood & Practical Deployment Guide*, McGraw-Hill, 2007, pogl. 1.
- [22] S. Viaene, »Linking Business Intelligence into Your Business,« *IT Pro*, November/December 2008, str. 28-34.
- [23] H. J. Watson, »Recent Development in Data Warehousing,« Communications of the Association for Information Systems, 2001.
- [24] H. J. Watson, B. H. Wixom, »The Current State of Business Intelligence,« *Computer*, September 2007, str. 96-99.
- [25] (2008) M. H. Weier. And Then There Were 4 – IBM, SAP, Oracle and Microsoft. TMCNet. Dostopno na: <http://www.tmcnet.com/usubmit/2008/04/14/3385297.htm>
- [26] (2006) Slovenski računovodski standardi, Ur.l. RS, št. 118/2005 (10/2006, 112/2006).
Dostopno na: <http://www.racunovodja.com/clanki.asp?clanek=631>